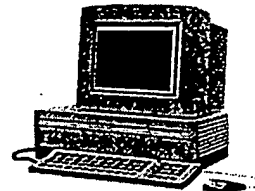


# BioTech-Chem Library

## Search Results

### Feedback Form (Optional)



Scientific & Technical Information C

The search results generated for your recent request are attached. If you have any questions or comments (compliments or complaints) about the scope or the results of the search, please contact *the BioTech-Chem searcher* who conducted the search *or contact*:

Mary Hale, Supervisor, 308-4  
CM-1 Room 1E01

---

#### *Voluntary Results Feedback Form*

➤ *I am an examiner in Workgroup:* (Example: 1610)

➤ *Relevant prior art found, search results used as follows:*

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

*Types of relevant prior art found:*

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ *Relevant prior art not found:*

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Search results were not useful in determining patentability or understanding the invention.

**Other Comments:**

---

Drop off completed forms at the Circulation Desk CM-1, or send to Mary Hale, CM1-1E01 or [mary.hale@uspto.gov](mailto:mary.hale@uspto.gov)

**THIS PAGE BLANK (USPTO)**

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

*The Pending database search results should not be left in the case because they contain data that is confidential.*

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



017N: 2d start

84524

Access DB#

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Celine Qian Examiner #: 75770 Date: 1/16/03  
Art Unit: 1636 Phone Number 306-0283 Serial Number: 09/822935  
Mail Box and Bldg/Room Location: 11E012 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Regulatory sequence of mouse villin gene - use in transgenics  
Inventors (please provide full names): Daniel Pinto et al.

Earliest Priority Filing Date: 2/19/1998

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please send for SEQ ID NO: 1 & a oligo search of SEQ ID NO.

CRFEE

Call  
2/1/03

RECEIVED  
JAN 16 2003  
STIC/Biotech

Edward Han  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

**THIS PAGE BLANK (USPTO)**



PN W020034492-A1.  
XX 15-JUN-2000.  
XX 09-DEC-1998; 98WO-EP08009.  
XX 09-DEC-1998; 98WO-EP08009.  
XX (CNRS ) CENT NAT RECH SCI.  
XX (CURIE-) INST CURIE.  
XX Plato D, Robine S, Jaisser F, Louvard D;  
XX WPI: 2000-423433/36.  
XX Novel nucleotide sequence derived from mouse villin gene for targeted  
PT expression of transgenes in immature and differentiated epithelial  
PT cells of intestine or urogenital tracts  
XX  
XX Claim 3; Fig 6; 54pp; English.  
XX The present sequence comprises the murine villin gene regulatory  
CC region, the first intron and part of the first two exons. It has been  
CC shown that this region directs the expression of the villin gene in the  
CC intestine and uro-genital tracts, and thus could be used in a fusion  
CC gene to direct expression of exogenous genes in these areas. This could  
CC be used, for example, to create a mouse model for colorectal cancer.  
XX  
SQ Sequence 8995 BP: 2275 A; 2105 C; 2258 G; 2357 T; 0 other:  
Query Match 100.0%; Score 8995; DB 21; Length 8995;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GATCGGTGACACAGACACTGTGTCCACACTGGGAGGTGAGAGGAGGTCA 60  
DB 1 GATCGGTGACACAGACACTGTGTCCACACTGGGAGGTGAGAGGAGGTCA 60  
QY 61 GAAGTTTAAAGTCATCTTGTGTACATAGCAAGTTTCAGCAGCTTACATGAAA 120  
DB 61 GAAGTTTAAAGTCATCTTGTGTACATAGCAAGTTTCAGCAGCTTACATGAAA 120  
QY 121 CTTTGTGTTGTTGTTGTTGTTTAAAGCATTAATAATTAATCAATAAGAGTTGG 180  
DB 121 CTTTGTGTTGTTGTTGTTGTTTAAAGCATTAATAATTAATCAATAAGAGTTGG 180  
QY 181 CAGTGTGGACAGACCTTAAATTCAGATATTCAGAGGACAGACAGATCTCTGT 240  
DB 181 CAGTGTGGACAGACCTTAAATTCAGATATTCAGAGGACAGACAGATCTCTGT 240  
QY 241 GAGTTCGAAGTCAGGCTGTGCAAGAGCTAGTTCAGAGTGGCAAGGCTTACACAGAGA 300  
DB 241 GAGTTCGAAGTCAGGCTGTGCAAGAGCTAGTTCAGAGTGGCAAGGCTTACACAGAGA 300  
QY 301 AACCTTGTCTCATAAACCAAGTAGTAGTAGTAATGTCATAGAGAAAATTGGA 360  
DB 301 AACCTTGTCTCATAAACCAAGTAGTAGTAGTAATGTCATAGAGAAAATTGGA 360  
QY 361 GTCCATTAGAGATGAGACCATCTATAGATGATTTCTTGACCCAGGTAAAGTAATGTCA 420  
DB 361 GTCCATTAGAGATGAGACCATCTATAGATGATTTCTTGACCCAGGTAAAGTAATGTCA 420  
QY 421 TGGGAAAGGGATGGAGCTGTCTAGATTAAAGGTCTGAGGGAGTCCATTCTCAA 480  
DB 421 TGGGAAAGGGATGGAGCTGTCTAGATTAAAGGTCTGAGGGAGTCCATTCTCAA 480  
QY 481 TTTGATTCATATGAAGAGCTGATAGAGCCCAAGAGAAGTGAAGTGGACTGTGACT 540  
DB 481 TTTGATTCATATGAAGAGCTGATAGAGCCCAAGAGAAGTGAAGTGGACTGTGACT 540  
QY 541 GAAGACGTGACGGCTTTAAACACTGGCACTTAAACACTTTAAACACTGGCACAGG 600  
DB 541 GAAGACGTGACGGCTTTAAACACTGGCACTTAAACACTTTAAACACTGGCACAGG 600

QY 601 CGTTCAGGTTTGAAGATCACTTTCAAAACACAGAGAAAGTGCGCTGCTCAGC 660  
DB 601 CGTTCAGGTTTGAAGATCACTTTCAAAACACAGAGAAAGTGCGCTGCTCAGC 660  
QY 661 GTAGGAGCACTGGCTGACAGAGAGTATTTAGTGAAGCTACCTTACAAATATCTTT 720  
DB 661 GTAGGAGCACTGGCTGACAGAGAGTATTTAGTGAAGCTACCTTACAAATATCTTT 720  
QY 721 GCACCTATACATATACAGTGTCAAAATGTCTACTCCCTAGTCCACAGATGGCTTTACA 780  
DB 721 GCACCTATACATATACAGTGTCAAAATGTCTACTCCCTAGTCCACAGATGGCTTTACA 780  
QY 781 CTCGTTTCTGCTTCCCATCTGTTGACATTTGTCAGAACCAAGAAATTTAGTGGCT 840  
DB 781 CTCGTTTCTGCTTCCCATCTGTTGACATTTGTCAGAACCAAGAAATTTAGTGGCT 840  
QY 841 ATTTATTTGTGTGCTGAGAGACACATCCAGGCTTTTACATTTTCAGGACATGCTTTAC 900  
DB 841 ATTTATTTGTGTGCTGAGAGACACATCCAGGCTTTTACATTTTCAGGACATGCTTTAC 900  
QY 901 TAACTGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTACTTATTTTGTCT 960  
DB 901 TAACTGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTACTTATTTTGTCT 960  
QY 961 GCATGAGGTAGGCAATGTATAGATGTATAGAGAGTCAATGTGAGCTGCTACCTCAAA 1020  
DB 961 GCATGAGGTAGGCAATGTATAGATGTATAGAGAGTCAATGTGAGCTGCTACCTCAAA 1020  
QY 1021 ATCATTTGCAGATCCCCACAGAGTGAAGTACAGGAGCTGTGTAGTTGTTATGTGGAGCTG 1080  
DB 1021 ATCATTTGCAGATCCCCACAGAGTGAAGTACAGGAGCTGTGTAGTTGTTATGTGGAGCTG 1080  
QY 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGACAGCAGTGGCTTTAACATGGACACAGCTCT 1140  
DB 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGACAGCAGTGGCTTTAACATGGACACAGCTCT 1140  
QY 1141 CTAGGCTTAAGATTAATCTTTGTTTAAATAATATATTTCTCAGCGGGGTGTGTCGC 1200  
DB 1141 CTAGGCTTAAGATTAATCTTTGTTTAAATAATATATTTCTCAGCGGGGTGTGTCGC 1200  
QY 1201 ACAGCCCTTTAATCCACAGCTTGAGAGGCTGAGAGTGTAGGAATTTATACACAGAGCCAG 1260  
DB 1201 ACAGCCCTTTAATCCACAGCTTGAGAGGCTGAGAGTGTAGGAATTTATACACAGAGCCAG 1260  
QY 1261 CTGGGCTGACAGAGCTTGGCCCTGTTTTTTTTTTTTTTTCTTATGTGACGTGTCTTA 1320  
DB 1261 CTGGGCTGACAGAGCTTGGCCCTGTTTTTTTTTTTTTTTCTTATGTGACGTGTCTTA 1320  
QY 1321 CCTGCGTATGTCCGTCGAAGGCTGTGAGATCCCTTGAGACTGGAGTTAAAGACAGTTG 1380  
DB 1321 CCTGCGTATGTCCGTCGAAGGCTGTGAGATCCCTTGAGACTGGAGTTAAAGACAGTTG 1380  
QY 1381 TGATCACGCTCCCGTTACAGATGCTGGAATTTGAACCCAGGCTGCTTAAGAGAACAGCC 1440  
DB 1381 TGATCACGCTCCCGTTACAGATGCTGGAATTTGAACCCAGGCTGCTTAAGAGAACAGCC 1440  
QY 1441 AGTGTCTTAATCTGTGAGCCACCCCTCAACCCGCTTTTGAAGACACTTAAACCTTTTG 1500  
DB 1441 AGTGTCTTAATCTGTGAGCCACCCCTCAACCCGCTTTTGAAGACACTTAAACCTTTTG 1500  
QY 1501 TGTAAATGTGGAAACTGAGTGTATCTTGACATTACCAAGTGTGTGCGCTGTAGACATCA 1560  
DB 1501 TGTAAATGTGGAAACTGAGTGTATCTTGACATTACCAAGTGTGTGTGCGCTGTAGACATCA 1560  
QY 1561 CTGAGCCCGTACCCACAGAGTGTGATAGAGTTTAAAGGCAACACTTAAACATGACA 1620  
DB 1561 CTGAGCCCGTACCCACAGAGTGTGATAGAGTTTAAAGGCAACACTTAAACATGACA 1620  
QY 1621 ATAGTGTGATAGAGTTTGAATATAGTCTGAGCTATTGTGTTAAGGCTTTTGTGTGCC 1680  
DB 1621 ATAGTGTGATAGAGTTTGAATATAGTCTGAGCTATTGTGTTAAGGCTTTTGTGTGCC 1680

OY	1681	TTACACATGCTGTGACAAACATGTGAACAAATGAAAGAACTTGGATGACTGTGCTCTGGAAACCCACA	1740
Db	1681	TTTACACATGCTGTGACAAACATGTGAACAAATGAAAGAACTTGGATGACTGTGCTCTGGAAACCCACA	1740
OY	1741	GAGCAGAGCCGAGAACCCACACTCTGAAGAGTTGTCTCTGAGCTTTCACATACAACTTCACAT	1800
Db	1741	GAGCAGAGCCGAGAACCCACACTCTGAAGAGTTGTCTCTGAGCTTTCACATACAACTTCACAT	1800
OY	1801	AATAGTTACAAATGATTAATTAATTAATAGTAAATCTTTTAAAGGTATATGTTGGAGGGA	1860
Db	1801	AATAGTTACAAATGATTAATTAATTAATAGTAAATCTTTTAAAGGTATATGTTGGAGGGA	1860
OY	1861	GAGATGGCTCAGCTTCCAGAGACACTTGGCTTCTGGCAGAGACTTGAATTCAGTTCC	1920
Db	1861	GAGATGGCTCAGCTTCCAGAGACACTTGGCTTCTGGCAGAGACTTGAATTCAGTTCC	1920
OY	1921	AGGACTATATATGAGGCTCTCACACCCATCTGTAAATCCATTCACAGAGTTCCACACCT	1980
Db	1921	AGGACTATATATGAGGCTCTCACACCCATCTGTAAATCCATTCACAGAGTTCCACACCT	1980
OY	1981	CCTTGGCCTTCACAGGACACACATACATAGTACACAGACATACATGACGCAAAACACC	2040
Db	1981	CCTTGGCCTTCACAGGACACACATACATAGTACACAGACATACATGACGCAAAACACC	2040
OY	2041	CATACACACATTAATTAATTAAGAAACCTTAAAGGTGCATGTGTTGGTAAACATTTGCT	2100
Db	2041	CATACACACATTAATTAATTAAGAAACCTTAAAGGTGCATGTGTTGGTAAACATTTGCT	2100
OY	2101	TACACATGCTGATTAAGAGACATGTACACAGGACACACCTAAAGGATCTGGGCGCTGAG	2160
Db	2101	TACACATGCTGATTAAGAGACATGTACACAGGACACACCTAAAGGATCTGGGCGCTGAG	2160
OY	2161	AGATGGCTCAGCGGTTAAAGACACTGACTGCTCTTCCGAGGAAAGTCTCTGAGTTCAAT	2220
Db	2161	AGATGGCTCAGCGGTTAAAGACACTGACTGCTCTTCCGAGGAAAGTCTCTGAGTTCAAT	2220
OY	2221	CCTAGCAACCATGCTGCTCTCAACACCATCCATTAATGAGATCTGACACCCCTCTTGCT	2280
Db	2221	CCTAGCAACCATGCTGCTCTCAACACCATCCATTAATGAGATCTGACACCCCTCTTGCT	2280
OY	2281	GCACTGAAAGACGCTCAGACGCTACAGTGTACTAGATATCTTAATTAATTAATCTTTT	2340
Db	2281	GCACTGAAAGACGCTCAGACGCTACAGTGTACTAGATATCTTAATTAATTAATCTTTT	2340
OY	2341	TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATGACAGTGCACG	2400
Db	2341	TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATGACAGTGCACG	2400
OY	2401	GGTATATATCTATCTGAGAGTTTCTTCTTCCGCTTGCTGCAACTGGGTGGACAGC	2460
Db	2401	GGTATATATCTATCTGAGAGTTTCTTCTTCCGCTTGCTGCAACTGGGTGGACAGC	2460
OY	2461	CCCTTTTCATTCACAAGAACGGGTCTACATTAATTTCTGAACCAAAACGACCTGCGAGT	2520
Db	2461	CCCTTTTCATTCACAAGAACGGGTCTACATTAATTTCTGAACCAAAACGACCTGCGAGT	2520
OY	2521	ATGTTTACTGTCTTGGTGCATATGAGACGGGACGCGCGGCGCACACACACACAC	2580
Db	2521	ATGTTTACTGTCTTGGTGCATATGAGACGGGACGCGCGGCGCACACACACACACAC	2580
OY	2581	ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGGTCA	2640
Db	2581	ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGGTCA	2640
OY	2641	AGAAAGAGGCTGCTCAAAACACATCTTCATCTTCCCTCTMAAGGAGACACAGATTCC	2700
Db	2641	AGAAAGAGGCTGCTCAAAACACATCTTCATCTTCCCTCTMAAGGAGACACAGATTCC	2700
OY	2701	AAGGTGGCAGAAATCTTACAGGGGGCAGAGGCGAGGGAGGGGAACGACGCAATGGTTTCC	2760
Db	2701	AAGGTGGCAGAAATCTTACAGGGGGCAGAGGCGAGGGAGGGGAACGACGCAATGGTTTCC	2760
OY	2761	AGAGACCTTACAGCAGAGGGGCGCAAGGACAGATCCCGAGGCTCCAGGGCAGGAGAGTGGAGG	2820

Db	2761	AGAGACCTACACGACGAGAGGACGACAGGACGATCCAGGCTCCAGGGACGGAGGGAGG	2820
QY	2821	CCCTGTTTCCGAGGACGAAGCAGGACGGCAGAACAGGGTTTCAAAAGCCACAGGTTTATGGCA	2880
Db	2821	CCCTTGTTCCGAGGACGAAGGACGGCGCAGAACAGGGTTCAAAAGCCACAGGTTTATGGCA	2880
QY	2881	GCTCATAAAGTGGAGGTCGTGCCTCACTCAGAAAGGAGGAAGAAAGGAGGAGGAGG	2940
Db	2881	GCTCATAAAGTGGAGGTCGTGCCTCACTCAGAAAGGAGGAAGAAAGGAGGAGGAGG	2940
QY	2941	GCCCACTGAGGAGGTCATGCTGATAGAGAGATCTGCAGAGGTCGCCAGGAGCCCCAC	3000
Db	2941	GCCCACTGAGGAGGTCATGCTGATAGAGAGATCTGCAGAGGTCGCCAGGAGCCCCAC	3000
QY	3001	CTGTCCTGCCAAGGGAAGCCCAAGTGTGAACCTCTGGCTTGAGTGTCCAGCTA	3060
Db	3001	CTGTCCTGCCAAGGGAAGCCCAAGTGTGAACCTCTGGCTTGAGTGTCCAGCTA	3060
QY	3121	CCCCGACTCCCGTGCACATTTCTTAGGGCTGAGAGGTCGCCAGCCCTGTGTGGGGTTGC	3180
Db	3121	CCCCGACTCCCGTGCACATTTCTTAGGGCTGAGAGGTCGCCAGCCCTGTGTGGGGTTGC	3180
QY	3181	CTACCTCAGGTAGAGCCCAAGGTCCTTAGCCGGGAAGTGCACCCCATCCCGAAGCTGCAGA	3240
Db	3181	CTACCTCAGGTAGAGCCCAAGGTCCTTAGCCGGGAAGTGCACCCCATCCCGAAGCTGCAGA	3240
QY	3241	GCCAAAGGGCGGGGCACACGGGACAGCTAGGCTGTGCAGAGCTGTGTGCTAGGCTTCC	3300
Db	3241	GCCAAAGGGCGGGGCACACGGGACAGCTAGGCTGTGCAGAGCTGTGTGCTAGGCTTCC	3300
QY	3301	AGGACCTGGGACCTACTTCCCAACCCCACTCCATTTCTCTGGGGCCCTATCTTCC	3360
Db	3301	AGGACCTGGGACCTACTTCCCAACCCCACTCCATTTCTCTGGGGCCCTATCTTCC	3360
QY	3361	CTTATATAGTGAAGGAGTTCGTGGGGGGGGGGGGGGTGGTGAAGCAAAAGTCGTTGC	3420
Db	3361	CTTATATAGTGAAGGAGTTCGTGGGGGGGGGGGGGGTGGTGGTGAAGCAAAAGTCGTTGC	3420
QY	3421	GTCCTCTGCACGGCAGCTTGCACACACTTCTTAAGATCTCCAGGTGTGGGCTCTTTC	3480
Db	3421	GTCCTCTGCACGGCAGCTTGCACACACTTCTTAAGATCTCCAGGTGTGGGCTCTTTC	3480
QY	3481	CAGACAGGTAAAGCAATTTGGGTGGGGACACATGTGTACACAGGTGTGGAGGGGACAG	3540
Db	3481	CAGACAGGTAAAGCAATTTGGGTGGGGACACATGTGTACACAGGTGTGGAGGGGACAG	3540
QY	3541	GGTCCTTGTCTTCTCTGCGAGGCTGTGCTTCTGTAGACACTTGTAAATTTGGGGG	3600
Db	3541	GGTCCTTGTCTTCTCTGCGAGGCTGTGCTTCTGTAGACACTTGTAAATTTGGGGG	3600
QY	3601	TGAGGTAAAGTGCTCTGAACCTCTGAAGAAGCAACAAAGCCAGAGGCTGTGGGCT	3660
Db	3601	TGAGGTAAAGTGCTCTGAACCTCTGAAGAAGCAACAAAGCCAGAGGCTGTGGGCT	3660
QY	3661	TCAATGAAGGAAGTTCACAGACCCCTTTCGTGAAGTCACCTTGCTTCATCTGTAG	3720
Db	3661	TCAATGAAGGAAGTTCACAGACCCCTTTCGTGAAGTCACCTTGCTTCATCTGTAG	3720
QY	3721	ATTCCCTGAGGACCAAGTGGCTCTGTGGACTCAGATTTTACAAATTAATCAGACAGT	3780
Db	3721	ATTCCCTGAGGACCAAGTGGCTCTGTGGACTCAGATTTTACAAATTAATCAGACAGT	3780
QY	3781	CCTGAGACTTGGACTCCGTGCTGTATTTACTACTTCTCTGTGGCTGCTCATTTCTGTGT	3840
Db	3781	CCTGAGACTTGGACTCCGTGCTGTATTTACTACTTCTCTGTGGCTGCTCATTTCTGTGT	3840
QY	3841	TCATGCTTACACATCTGAATGGTTTCTTTGTGTACCATTCCTCCCTACACTCTCTGGGA	3900

Db 3841 TCATGCTTACACATCTGAAATGTTTCTTGTGTACCATTCCTCCGACACATCCTGGGA 3900  
Oy 3901 GGTGCTATCCTGGACATGTATCCTGGGATGTAGCTGCAGCCACCGAGAGAGGGGG 3960  
Db 3901 GGTGCTATCCTGGACATGTATCCTGGGATGTAGCTGCAGCCACCGAGAGAGGGGG 3960  
Oy 3961 AGAGTCAGAGAGCTGTCTCTAGGCCCCATTTAGGCCCCGACATCACCCCTTCTCTAGAAAT 4020  
Db 3961 AGAGTCAGAGAGCTGTCTCTAGGCCCCATTTAGGCCCCGACATCACCCCTTCTCTAGAAAT 4020  
Oy 4021 GGGCCCCCTCCATTTTCGGTTACCATGTATTTATATCAGAGTGGGAGGAAAGCCA 4080  
Db 4021 GGGCCCCCTCCATTTTCGGTTACCATGTATTTATATCAGAGTGGGAGGAAAGCCA 4080  
Oy 4081 AACCTGCCACGAAAGTTTGGGACTCAGACCAAGGTTATGTCTCGAAATCCCCCTG 4140  
Db 4081 AACCTGCCACGAAAGTTTGGGACTCAGACCAAGGTTATGTCTCGAAATCCCCCTG 4140  
Oy 4141 TCACCTTGAGTTGGGAGATCTGCTCTGGGGCTTCAGAGCTTGGTTAGCAGAGAGGT 4200  
Db 4141 TCACCTTGAGTTGGGAGATCTGCTCTGGGGCTTCAGAGCTTGGTTAGCAGAGAGGT 4200  
Oy 4201 ATCTTTGTATATAGGCATGACCTAGTCTATGTGTGTACTACTTCTCTGCACTTAAAG 4260  
Db 4201 ATCTTTGTATATAGGCATGACCTAGTCTATGTGTGTACTACTTCTCTGCACTTAAAG 4260  
Oy 4261 CTGGAAGTAAACCCACGCGACGCGCCAGGATTTCTACAGTTGTAACCCCAAGCAACA 4320  
Db 4261 CTGGAAGTAAACCCACGCGACGCGCCAGGATTTCTACAGTTGTAACCCCAAGCAACA 4320  
Oy 4321 AGACAGTATATGCAAGATAGTAGTGGGGAGAGAACTTAAACCCCCCAAG 4380  
Db 4321 AGACAGTATATGCAAGATAGTAGTGGGGAGAGAACTTAAACCCCCCAAG 4380  
Oy 4381 GGGCAGAGTTCCCTTCCCTACTTACCAATGCGACATATAGTGTAGTACTACTATAGGCTG 4440  
Db 4381 GGGCAGAGTTCCCTTCCCTACTTACCAATGCGACATATAGTGTAGTACTACTATAGGCTG 4440  
Oy 4441 TGAGTTGGAGCTACAGCATGATGATGTTTCATGTGTGTAGTGTATATATCTAGACAC 4500  
Db 4441 TGAGTTGGAGCTACAGCATGATGATGTTTCATGTGTGTGTGTGTATATCTAGACAC 4500  
Oy 4501 TTGGAGGCTGAGACGAGAGATGCTATATGTTTGAAGCCAGCTGAGCTTATAGCGA 4560  
Db 4501 TTGGAGGCTGAGACGAGAGATGCTATATGTTTGAAGCCAGCTGAGCTTATAGCGA 4560  
Oy 4561 GACTTGTCTTTAAGAAAAAATGAAGCCCGACAGTGTGGCAACGCCCTTATCCCA 4620  
Db 4561 GACTTGTCTTTAAGAAAAAATGAAGCCCGACAGTGTGGCAACGCCCTTATCCCA 4620  
Oy 4621 GCACCTTGGAGGACAGACGAGATTTCTGATTTCAAGGCGACGCTGTCTATAGAGT 4680  
Db 4621 GCACCTTGGAGGACAGACGAGATTTCTGATTTCAAGGCGACGCTGTCTATAGAGT 4680  
Oy 4681 GAGTTTCCAGGACAGCCAGGGCTACACAGAGAAACCTGTTTGAAGAAACCAAGAAACAA 4740  
Db 4681 GAGTTTCCAGGACAGCCAGGGCTACACAGAGAAACCTGTTTGAAGAAACCAAGAAACAA 4740  
Oy 4741 AACCAAAACAAAAACAAACCAACCAACCAACCAACCAACCTCTCATCTCTCATCTCTC 4800  
Db 4741 AACCAAAACAAAAACAAACCAACCAACCAACCAACCAACCTCTCATCTCTCTC 4800  
Oy 4801 TAGGCTGTCTGTCTAGTGTGTAGTGTGTGGGACTTCAGACTTATATATATATAGGCC 4860  
Db 4801 TAGGCTGTCTGTCTAGTGTGTAGTGTGTGGGACTTCAGACTTATATATATATAGGCC 4860  
Oy 4861 TTTTATACATGTGTGACAGAGAGAAAGTTTCACTGTGGACACAGTGGACCTTGAGA 4920  
Db 4861 TTTTATACATGTGTGACAGAGAGAAAGTTTCACTGTGGACACAGTGGACCTTGAGA 4920  
Oy 4921 AAGTACTCTTGGCAGCCCAAAATTTGGGAAGGCTTCTGTGAGAGAGTGTCTCCGAT 4980  
Db 4921 AAGTACTCTTGGCAGCCCAAAATTTGGGAAGGCTTCTGTGAGAGAGTGTCTCCGAT 4980

Oy 4981 CAGACTACTGTCTAGAGGACAGAGAGAGGGTTGGAAGATGTTGTTGGACAGACACTT 5040  
Db 4981 CAGACTACTGTCTAGAGGACAGAGAGAGGGTTGGAAGATGTTGTTGGACAGACACTT 5040  
Oy 5041 GGAACAGAGGACAGAGAGGGGAGGCATCCAGATTTCTGAACATGTAGCTGACTTTTGGT 5100  
Db 5041 GGAACAGAGGACAGAGAGGGGAGGCATCCAGATTTCTGAACATGTAGCTGACTTTTGGT 5100  
Oy 5101 TCTCTGGGTGACAGTGTCCCCAGGGATAGGGCTGTAGAAAGGGGACCCAGGGGTGACC 5160  
Db 5101 TCTCTGGGTGACAGTGTCCCCAGGGATAGGGCTGTAGAAAGGGGACCCAGGGGTGACC 5160  
Oy 5161 AATGAGTTCAGTTGAGGAGACATCCAGCCAGGCTGCTGTGGCAAGCTTAAAGATG 5220  
Db 5161 AATGAGTTCAGTTGAGGAGACATCCAGCCAGGCTGCTGTGGCAAGCTTAAAGATG 5220  
Oy 5221 AGAGCCCTCTAACCCCTCCCTGAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTCTA 5280  
Db 5221 AGAGCCCTCTAACCCCTCCCTGAAGTTTAGGGAGAGACAGAGAGCTGAGAGATCCTCTA 5280  
Oy 5281 GGTGAAGAGAGATGTGCTGTGACCAACATGGCTTAGAGACAGACAGTGTGACAG 5340  
Db 5281 GGTGAAGAGAGATGTGCTGTGACCAACATGGCTTAGAGACAGACAGTGTGACAG 5340  
Oy 5341 TTACCCCTCAGAACCCAGCCATCCCTTGGCTTAAGGAGGCTGGGCCCTTCTGTTT 5400  
Db 5341 TTACCCCTCAGAACCCAGCCATCCCTTGGCTTAAGGAGGCTGGGCCCTTCTGTTT 5400  
Oy 5401 AAGATCTTACTTTTCTTACAGAGAGGACAGACCCCTTGTCCCTCCTGTTGGTCAA 5460  
Db 5401 AAGATCTTACTTTTCTTACAGAGAGGACAGACCCCTTGTCCCTCCTGTTGGTCAA 5460  
Oy 5461 TAAACCCCTGTGTGTAACATTAATTTACTGTGACAGTTGCTCAGAGACGTCCA 5520  
Db 5461 TAAACCCCTGTGTGTAACATTAATTTACTGTGACAGTTGCTCAGAGACGTCCA 5520  
Oy 5521 TCTGTAGACCTGTGCTCTTAACCTCACCAGGATAGGCCCATTTCTCACCAGAGAG 5580  
Db 5521 TCTGTAGACCTGTGCTCTTAACCTCACCAGGATAGGCCCATTTCTCACCAGAGAG 5580  
Oy 5581 TGCAGAGAGAGCCTTACAGAAAGGGTAAACAGTAAACAAAGATGGCCAGATTAACAA 5640  
Db 5581 TGCAGAGAGAGCCTTACAGAAAGGGTAAACAGTAAACAAAGATGGCCAGATTAACAA 5640  
Oy 5641 ACTATATCCTTTGTAACCAATTTGGTTTGTGTAACAGAGAGGGGTGTGAGTGTAT 5700  
Db 5641 ACTATATCCTTTGTAACCAATTTGGTTTGTGTAACAGAGAGGGGTGTGAGTGTAT 5700  
Oy 5701 GT 5760  
Db 5701 GT 5760  
Oy 5761 CTTGGGGACCTTTTCATCTTAAGAAATATCTGATTTTGGCCCCATGCAACAGGGGTAT 5820  
Db 5761 CTTGGGGACCTTTTCATCTTAAGAAATATCTGATTTTGGCCCCATGCAACAGGGGTAT 5820  
Oy 5821 TGGGAGAGTCAAGCTTGTGCAACACAGTAACTGCGCCCAAGATGATGTGGGCTAA 5880  
Db 5821 TGGGAGAGTCAAGCTTGTGCAACACAGTAACTGCGCCCAAGATGATGTGGGCTAA 5880  
Oy 5881 TCACCAAGGAGGAGCTGATCAGATGAGACAGAACATCACAAGATTAACCAACCTGTGG 5940  
Db 5881 TCACCAAGGAGGAGCTGATCAGATGAGACAGAACATCACAAGATTAACCAACCTGTGG 5940  
Oy 5941 GCTCAGAGAGGAGCTTACAGAGAGTAAAGGCCAAGCCATTTATTTACAGACATGAC 6000  
Db 5941 GCTCAGAGAGGAGCTTACAGAGAGTAAAGGCCAAGCCATTTATTTACAGACATGAC 6000  
Oy 6001 TCAAAATCAAAAGTGCAGAGAGATTAAGTGTGAGAGATGGGGGTGTGAGTGTGGACACC 6060  
Db 6001 TCAAAATCAAAAGTGCAGAGAGATTAAGTGTGAGAGATGGGGGTGTGAGTGTGGACACC 6060

Oy	6061	TCACCTTGCACTTATTAGTACTAGGCCAAAGAGCAGCTACAGAGGGTACTGGGCTCTA	6120
Oy	6061	TCACCTTGCACTTATTAGTACTAGGCCAAAGAGCAGCTACAGAGGGTACTGGGCTCTA	6120
Oy	6121	CTCAGCTTGGAGCGAGCAGCTGGAGAAATGGGTGACCCTCATCTCATGTAGAGGGGCTAG	6180
Oy	6121	CTCAGCTTGGAGCGAGCAGCTGGAGAAATGGGTGACCCTCATCTCATGTAGAGGGGCTAG	6180
Oy	6181	CACGACAGGATACAGAGTGTCCCTGTGTCTCATGTCCAGGATTCCTGGCAGTTTTCAAAG	6240
Oy	6181	CACGACAGGATACAGAGTGTCCCTGTGTGTCTCATGTCCAGGATTCCTGGCAGTTTTCAAAG	6240
Oy	6241	GACTAAGGACTCATCTCTGGTGGAAACAAAGTATCCAGCCTTAAGCCCCATTTTGGTCT	6300
Oy	6241	GACTAAGGACTCATCTCTGGTGGAAACAAAGTATCCAGCCTTAAGCCCCATTTTGGTCT	6300
Oy	6301	AATTAATCAGAAACCCCTGGGGATGCGAGGCTCGAGCAGCAGACACTTTTAAAAAGTC	6360
Oy	6301	AATTAATCAGAAACCCCTGGGGATGCGAGGCTCGAGCAGCAGACACTTTTAAAAAGTC	6360
Oy	6361	CCAGGTATTTCTGATCAGCAGCTGGAAACAAACAGACTACAGTTTCAACAGAAAGAGC	6420
Oy	6361	CCAGGTATTTCTGATCAGCAGCTGGAAACAAACAGACTACAGTTTCAACAGAAAGAGC	6420
Oy	6421	AAACCTTAGGGAACCTTGGGATGGGAGGCTTCTTCAGGCGCAGTATGAGAGCTGTT	6480
Oy	6421	AAACCTTAGGGAACCTTGGGATGGGAGGCTTCTTCAGGCGCAGTATGAGAGCTGTT	6480
Oy	6481	AGCAGTGGTGGACACTCTCTCTGCGCTGTCATATACCTATCCATCCATCCATCCAT	6540
Oy	6481	AGCAGTGGTGGACACTCTCTCTGCGCTGTCATATACCTATCCATCCATCCATCCAT	6540
Oy	6541	ACACCCACCCATCCATTTATGCAACCCATCCCTTCATCCATCCATCCATCCATCCATCC	6600
Oy	6541	ACACCCACCCATCCATTTATGCAACCCATCCCTTCATCCATCCATCCATCCATCCATCC	6600
Oy	6601	CCACGACATCCATCCAAACCTTCTTTCCTCTCTTCTTCTTCTTCTTCTTCTTCTTCT	6660
Oy	6601	CCACGACATCCATCCAAACCTTCTTTCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	6660
Oy	6661	CATTATATCCAAAGAGAACTGATATGATGATGATGATGATGATGATGATGATGATGAT	6720
Oy	6661	CATTATATCCAAAGAGAACTGATATGATGATGATGATGATGATGATGATGATGATGAT	6720
Oy	6721	AGCTCTGTGATGACTGATTTGATGATGATGATGATGATGATGATGATGATGATGATG	6780
Oy	6721	AGCTCTGTGATGACTGATTTGATGATGATGATGATGATGATGATGATGATGATGATG	6780
Oy	6781	TGGCAATCGGAGAAAGGTTTTGGGTGTTGTTTTCTCTCCACCGTGGGTCTGGGGA	6840
Oy	6781	TGGCAATCGGAGAAAGGTTTTGGGTGTTGTTTTCTCTCCACCGTGGGTCTGGGGA	6840
Oy	6841	TTGGAATCCAAATTAATCGGCGTGGGTGGCAGTGTCTTACACCGAGCCATTTTGTGACA	6900
Oy	6841	TTGGAATCCAAATTAATCGGCGTGGGTGGCAGTGTCTTACACCGAGCCATTTTGTGACA	6900
Oy	6901	CATCATTTATTATTAAGAAGACTCTTATGATGATGATGATGATGATGATGATGATGAT	6960
Oy	6901	CATCATTTATTATTAAGAAGACTCTTATGATGATGATGATGATGATGATGATGATGAT	6960
Oy	6961	CCACGATGACCTTAACTCTCTCTTCAGCCTCCACCCAGATGCTAGGTTTACAGT	7020
Oy	6961	CCACGATGACCTTAACTCTCTCTTCAGCCTCCACCCAGATGCTAGGTTTACAGT	7020
Oy	7021	GTTCAACTGGTGAATGCCCTTAAATCCAGACTCTGTGGGGGGGGGGGGGGGGGGATC	7080
Oy	7021	GTTCAACTGGTGAATGCCCTTAAATCCAGACTCTGTGGGGGGGGGGGGGGGGGGATC	7080
Oy	7081	CCTGAGTTGGAGGCACTTTTGTGCTCAGAGTTTCAGGATACCTGGGCTATACAGGGA	7140
Oy	7081	CCTGAGTTGGAGGCACTTTTGTGCTCAGAGTTTCAGGATACCTGGGCTATACAGGGA	7140
Oy	7141	ACCCTATCCCAACAAACAAACAAACAAACAAATATTTGTCTCAATATATCAGAGA	7200

[illegible]

Db	8221	TGATTATATGATGTTTATTTTAAAGTGGCCCAAGATACAGAGTATTTTGTGGAGTTTTCAC	8280
Oy	8281	CTTCCCTTTGAGGCTCTCGCATTTAACTACGTCCCTGGGCTAGTGA6CAATGCTTCA	8340
Db	8281	CTTCCCTTTGAGGCTCTCGCATTTAACTACGTCCCTGGGCTAGTGA6CAATGCTTCA	8340
Oy	8341	CTGCATGAGGCATCTGGCTGCTGCTGGCAAGCTCCCTCTATTTCACAGATGGGACTA	8400
Db	8341	CTGCATGAGGCATCTGGCTGCTGCTGGCAAGCTCCCTCTATTTCACAGATGGGACTA	8400
Oy	8401	CGCACTGCATGAGCCTTAAAGCTCAGCAAGTCAATCAGAGTGGCTAGCAGGAGACTCAG	8460
Db	8401	CGCACTGCATGAGCCTTAAAGCTCAGCAAGTCAATCAGAGTGGCTAGCAGGAGACTCAG	8460
Oy	8461	CGATATGCTGGGCTCTGCTCCACAGTGGTACGATTTACAGGCATACATCTACTGCTGAG	8520
Db	8461	CGATATGCTGGGCTCTGCTCCACAGTGGTACGATTTACAGGCATACATCTACTGCTGAG	8520
Oy	8521	ATTTTAAACCGAATCCTGAGATAGAGCAGGACTCTACAAATGAGAGGTTCTTTTGT	8580
Db	8521	ATTTTAAACCGAATCCTGAGATAGAGCAGGACTCTACAAATGAGAGGTTCTTTTGT	8580
Oy	8581	GTTTGGTTTGGTTTCTCTGCTATTAAGATCAGCAGTCTGAAATAGTATAGCTGGGCTAC	8640
Db	8581	GTTTGGTTTGGTTTCTCTGCTATTAAGATCAGCAGTCTGAAATAGTATAGCTGGGCTAC	8640
Oy	8641	ATTAACATCTTGTCTCAAAAAGCCCTATAGAGGTGAGGAGTGCAGGCTTAAAGAAGCCTT	8700
Db	8641	ATTAACATCTTGTCTCTCAAAAAGCCCTATAGAGGTGAGGAGTGCAGGCTTAAAGAAGCCTT	8700
Oy	8701	AAGCGGCTGTATAGCACACAGAGATAGCCTGCATATATAGCAAGACCTTGTTCAAA	8760
Db	8701	AAGCGGCTGTATAGCACACAGAGATAGCCTGCATATATAGCAAGACCTTGTTCAAA	8760
Oy	8761	ACATGAGAGGAGGGGTATGTTTAACTGCTGGGCTGTGTAAGGCACCTAAGAGGCCAA	8820
Db	8761	ACATGAGAGGAGGGGTATGTTTAACTGCTGGGCTGTGTAAGGCACCTAAGAGGCCAA	8820
Oy	8821	TGTAGACATTTGACTAAGAAAGATCATCATCAAAAGCCGGTGGGAGGATAGAGTTGG	8880
Db	8821	TGTAGACATTTGACTAAGAAAGATCATCATCAAAAGCCGGTGGGAGGATAGAGTTGG	8880
Oy	8881	ACTACAGTGGTCAAGAACCCCATTAAGAGCAAGTTCCTTCTCTGAGGCTCAAGC	8940
Db	8881	ACTACAGTGGTCAAGAACCCCATTAAGAGCAAGTTCCTTCTCTGAGGCTCAAGC	8940
Oy	8941	CTGAGCTGCAGGCGACACTGCTCTCAATGCCCTTCTCTAGGCTGTCACCATG	8995
Db	8941	CTGAGCTGCAGGCGACACTGCTCTCAATGCCCTTCTCTAGGCTGTCACCATG	8995
RESULT 2			
AAAA9605	AAAA9605 standard; DNA; 8993 BP.		
XX	AAAA9605:		
XX	AC		
DT	19-DEC-2000 (first entry)		
XX			
DE	Mouse villin gene regulatory elements and partial sequence.		
XX			
KW	Mouse: villin; intestinal epithelial cell;		
XX	uro-genital tract epithelial cell; tumour; ds.		
OS	Mus sp.		
XX			
FH	Key	Location/Qualifiers	
FT	misc_signal	3442	
FT		/*tag= a	
FT	exon	/note="transcriptional start site"	
FT		3442..3487	
FT		/*tag= b	
FT		/number= 1	

FT	intron	3488..8981
FT	/tag= c	
FT	/number= 1	
FT	exon	8982..8995
FT	/tag= d	
FT	/number= 2	
FT	misc_signal	8992
FT	/*tag= e	
FT	/note= "translational start site"	
PN		
PX	WO200034493-AZ.	
PD	15-JUN-2000.	
PX		
PF	09-DEC-1999;	99WO-EP09782.
PX		
PR	09-DEC-1998;	98WO-EP08009.
PX		
PA	(CNRS ) CENT NAT RECH SCI.	
PA	(CURIE-) INST CURIE.	
PI		
PI	Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J,	
DR	WPI; 2000-423434/36.	
PX		
PT	Novel nucleotide sequence derived from mouse villin gene for targeted expression of transgenes in immature and differentiated epithelial cells of intestine or urogenital tracts -	
PS	Claim 3; Fig 6; 52pp; English.	
PX		
CC	The present sequence consists of the regulatory region and the first exon and intron of the murine villin gene. This gene is expressed in the epithelial cells of the intestine and uro-genital tracts. Its promoter sequence can be used in the targeted expression of exogenous genes in these places, which may, for example, be useful in the treatment of tumours.	
CC		
CC		
SQ	Sequence 8993 BP: 2274 A; 2105 C; 2257 G; 2355 T; 2 other:	
Query Match            96.2%; Score 8651; DB 21; Length 8993;		
Best Local Similarity   100.0%; Pred. NO. 0;		
Matches 8991; Conservative   0; Mismatches   2; Indels   2; Gaps   2.		
OY	1 GATCTGTCGTACACAAGGACACTGTGGTCCCACACTGGGGAGTGAGGAGAGGCGTCA 60	
DB	1 GATCTGTCGTACACAAGGACACTGTGGTCCCACACTGGGGAGTGAGGAGAGGCGTCA 60	
OY	61 GAAGTTAAGTCATCCTTGTTTACATACGAAGTTTCAGCCAGCTTCAAGCATGAAA 120	
DB	61 GAAGTTAAGTCATCCTTGTTTACATACGAAGTTTCAGCCAGCTTCAAGCATGAAA 120	
OY	121 CCTTGTTGTTGTTGTTGTTGTTTAAAGCATTAATAATATACATAAGAGGATTGG 180	
DB	121 CCTTGTTGTTGTTGTTGTTGTTTAAAGCATTAATAATATACATAAGAGGATTGG 180	
OY	181 CAGTGTGCAGACACCCTTAATTCACATATTTCAGAGGACAGAAGCAGATCTCTGT 240	
DB	181 CAGTGTGCAGACACCCTTAATTCACATATTTCAGAGGACAGAAGCAGATCTCTGT 240	
OY	241 GAGTTCGAAGTAGGCTATCTGTGCCAAGCTATTCAGAGTAGGCAAGGCTCACACAGA 300	
DB	241 GAGTTCGAAGTAGGCTATCTGTGCCAAGCTATTCAGAGTAGGCAAGGCTCACACAGA 300	
OY	301 AACCTGTCTCTAAAAAACCAAAGTAGTACTAGTAGTAATGCCATAGAAAAATTGGA 360	
DB	301 AACCTGTCTCTAAAAAACCAAAGTAGTACTAGTAGTAATGCCATAGAAAAATTGGA 360	
OY	361 GTTCATATCGAGTATGAGCATCTCTATTAAGATGATTCTCTTGACCCAGGTAAAGTATGTA 420	
DB	361 GTTCATATCGAGTATGAGCATCTCTATTAAGATGATTCTCTTGACCCAGGTAAAGTATGTA 420	
OY	421 TGGGGAAAGGGAGTGGACTGTCTAGATTAAAAAGTGTGAGGGGATGCCATTCTCAA 480	





QY	264.1	AGAAAGAGGCTCCCTCAACACAGATCTTCATCTTTCCCTCTTAANGAGACCAGATTC	2700
Db	264.1	AGAAAGAGGCTCCCTCCCAACACAGATCTTCATCTTTCCCTCTTAANGAGAGACAGATTC	2700.0
QY	270.1	AAGGAGGAGAAAGTCTACAGAGGGGGCAGAGGCGAGGGAGGGGAGACAGCCATGGTTTCC	2760
Db	270.1	AAGGTGGCAGAAAGTCTACAGAGGGGGCAGAGGCGAGGGAGGGGAGACAGCCATGGTTTCC	2760.0
QY	276.1	AGAGACTACAGCAGAGGGGACGAGCAAGGATCCAGGTCCAGGGCAGGGAGGTGAGG	2820
Db	276.1	AGAGAGCTACAGCAGAGGGGACGAGCAAGGATCCAGGTCCAGGGGACGGAGGTGAGG	2820.0
QY	282.1	CCCTTGTTCCGAGGAGAAAGGCGAGGGCGAGAACAGGGTTCAAAAGGCACAGGTTTATGCA	2880
Db	282.1	CCCTTGTTCCGAGGAGAAAGGCGAGGGCGAGAACAGGGTTCAAAAGGCACAGGTTTATGCA	2880.0
QY	288.1	GCTCATAAAGTGAGAGGTCCGTGGCTCACTACAGAAAGAGAAAGAGGAAAGGCCCTTGT	2940
Db	288.1	GCTCATAAAGTGAGAGGTCCGTGGCTCACTACAGAAAGAGAAAGAGGAAAGGCCCTTGT	2940.0
QY	294.1	GCCCACTGAGGAGAGGTCTAGTAGAGAGATCTCAGAGGGTGGCAGAGAGCCCA	3000
Db	294.1	GCCCACTGAGGAGAGGTCTAGTAGAGAGATCTCAGAGGGTGGCAGAGAGCCCA	3000.0
QY	300.1	CTGTCTGTCCCAAGGAAACCCTCAAGTGTAACCTGTGGCTGGAGTTCAGACTA	3060
Db	300.1	CTGTCTGTCCCAAGGAAACCCTCAAGTGTAACCTGTGGCTGGAGTTCAGACTA	3060.0
QY	306.1	CNAGACCCCAAGAGTCTTACTCATCCATCCCATAGTGCCCTCGCCCGCCACACCCCA	3120
Db	306.1	CNAGACCCCAAGAGTCTTACTCATCCATCCCATAGTGCCCTCGCCCGCCACACCCCA	3120.0
QY	312.1	CCCCGACCTCCCGGACACTTCCTAGAGGCTGGAGGTCGAGGCGGAGGTTGGC	3180
Db	312.1	CCCCGACCTCCCGGACACTTCCTAGAGGCTGGAGGTCGAGGCGGAGGTTGGC	3180.0
QY	318.1	CTACCTGCAAGTAGAGCCAGGTCCTAGGCGGAAGTGACACCCCATCCCTGAAGCTCAGA	3240
Db	318.1	CTACCTGCAAGTAGAGCCAGGTCCTAGGCGGAAGTGACACCCCATCCCTGAAGCTCAGA	3240.0
QY	324.1	GCCAAAGGGCGGGGACACAGGCGACTCAAGGCTGTGACGGCTGTGGGCTCTAGATTC	3300
Db	324.1	GCCAAAGGGCGGGGACACAGGCGACTCAAGGCTGTGACGGCTGTGGGCTCTAGATTC	3300.0
QY	330.1	AGGACCTGGGGCAGCTACTTCCACCACCCCCCATTCATCTCTCGGGGCCCTATCTTC	3360
Db	330.1	AGGACCTGGGGCAGCTACTTCCACCACCCCCCATTCATCTCTCGGGGCCCTATCTTC	3360.0
QY	336.1	CTTATATGTGTGAAGAGTTCCTGGGGGGGGGGGTGGTGTGAGGACAAAGTCTTTCG	3420
Db	336.1	CTTATATGTGTGAAGAGTTCCTGGGGGGGGGGGTGGTGTGAGGACAAAGTCTTTCG	3420.0
QY	342.1	GTTCTCTGACAGCCAGCTTGGGCACAACTCTTCAAGATCTCCAGGTGTGGCTCTTC	3480
Db	342.1	GTTCTCTGACAGCCAGCTTGGGCACAACTCTTCAAGATCTCCAGGTGTGGCTCTTC	3480.0
QY	348.1	CAGACAGGTAAAGGCAATTTGGGTGGGACACATGTGACCCACAGGTGTGGAGGGAG	3540
Db	348.1	CAGACAGGTAAAGGCAATTTGGGTGGGACACATGTGACCCACAGGTGTGGAGGGAG	3540.0
QY	354.1	GGTCCCTGGCTCTCTCTGGAGAGCTGGCTTCTGAGACCTTGGTAAAGTTTGGGG	3600
Db	354.1	GGTCCCTGGCTCTCTCTGGAGAGCTGGCTTCTGAGACCTTGGTAAAGTTTGGGG	3600.0
QY	360.1	TGAGTAAAGTGCTCTGAAACTCTGAAAGAGCAGAAACCCAGCAGGCTGTGGGGCT	3660
Db	360.1	TGAGTAAAGTGCTCTGAAACTCTGAAAGAGCAGAAACCCAGCAGGCTGTGGGGCT	3660.0
QY	366.1	TCATATGAAGGAATTCACACAGCCCTTTCCTGTAGTACCTTGCTTACATCTGTGAG	3720
Db	366.1	TCATATGAAGGAATTCACACAGCCCTTTCCTGTAGTACCTTGCTTACATCTGTGAG	3720.0

QY	3721	ATTCCCTGGAAACCAAGTGGCTCCTGGGACTCAGATTTTCTACAATTAAATTCAGACAGT	3780
Db	3721	ATTCCTCTGGGACCAAGGTGGCTCCTGGGACTCAGATTTTCTACAATTAAATTCAGACAGT	3780
QY	3781	CCTGAGACTTGGACTCCGTCGCTGTATTATACATCTCTCTGCGCGTCATATTCTGTGT	3840
Db	3781	CCTGAGACTTGGACTCCGTCGCTGTATTATACATCTCTCTGCGCGTCATATTCTGTGT	3840
QY	3841	TCATGTCTTACACATCTGAAATGGTTTCTTGTGTGTGCACATTCCCTGACACTCCTGGGA	3900
Db	3841	TCATGTCTTACACATCTGAAATGGTTTCTTGTGTGTGCACATTCCCTGACACTCCTGGGA	3900
QY	3901	GGTGTATCTTGGCACATGATATCTCGGGATGTAGCTCAGCCACACAGAGAGAGGGG	3960
Db	3901	GGTGTATCTTGGCACATGATATCTCGGGATGTAGCTCAGCCACACAGAGAGAGGGG	3960
QY	3961	AGAGCAGAGACTGTGTCTTAGGCGCATTTAGCGCCAGAGATCACCCCTTTCCTAGAAAT	4020
Db	3961	AGAGCAGAGACTGTGTCTTAGGCGCATTTAGCGCCAGAGATCACCCCTTTCCTAGAAAT	4020
QY	4021	GGCCCTTCATTTTTCCGTTTACCATGATCATTTTATACAGAGTGGGAGTGAAGGCA	4080
Db	4021	GGCCCTTCATTTTTCCGTTTACCATGATCATTTTATACAGAGTGGGAGTGAAGGCA	4080
QY	4081	AACCTGCCAGAGTTTGGACATCTACTCAGACCAGGTTATCTGCTCAGAAATCCCTGT	4140
Db	4081	AACCTGCCAGAAATTTTGGGACTCAGACAGCCAAAGTTATCTGCTCAGAAATCCCTGT	4140
QY	4141	TCACCTGAGGTTGGGAAATCTGCTCTGGGGGCTTCAGAGTCTTGTTAGCAGAGGGT	4200
Db	4141	TCACCTGAGGTTGGGAAATCTGCTCTGGGGGCTTCAGAGTCTTGTTAGCAGAGGGT	4200
QY	4201	ATTCCTTTATAGGGGATGTGACTAGTCTATAGTGTACTACATTCCTGTCCAGTTTAAAG	4260
Db	4201	ATTCCTTTATAGGGGATGTGACTAGTCTATAGTGTACTACATTCCTGTCCAGTTTAAAG	4260
QY	4261	CTGGACTTAAACCCACAGCGAGCGCCAGAGATTCTCTACAGTTTGAACCCCAAGACA	4320
Db	4261	CTGGACTTAAACCCACAGCGAGCGCCAGAGATTCTCTACAGTTTGAACCCCAAGACA	4320
QY	4321	AGACAGTAGATATCCAGAGATAGGTAGCTGGGGAAGAACTTAAACCCCCCAAG	4380
Db	4321	AGACAGTAGATATCCAGAGATAGGTAGCTGGGGAAGAACTTAAACCCCCCAAG	4380
QY	4381	GCCCAAGTTCCTTCCCTAGTTTCAATATGCCAGTATAGTGTAGCTATATGAGCTG	4440
Db	4381	GCCCAAGTTCCTTCCCTAGTTTCAATATGCCAGTATAGTGTAGCTATATGAGCTG	4440
QY	4441	TGACTTGGTACTACAGCATGATGATGTTCATGTGTGTAGTGTGTATATCTGAGCAC	4500
Db	4441	TGACTTGGTACTACAGCATGATGATGTTCATGTGTGTAGTGTGTATATCTGAGCAC	4500
QY	4501	TTGGGAGGCTAAGCAGAGAGATTGCTATATGTTTAGGCCACGCTGAGCATATAGAGCA	4560
Db	4501	TTGGGAGGCTAAGCAGAGAGATTGCTATATGTTTAGGCCACGCTGAGCATATAGAGCA	4560
QY	4561	GACTTGTCTTTAGAAAAATGAAAGCCACAGTGTGGCACACAGCCTTTTATATCCA	4620
Db	4561	GACTTGTCTTTAGAAAAATGAAAGCCACAGTGTGGCACACAGCCTTTTATATCCA	4620
QY	4621	GCACCTTGGAGGAGAAAGCAGGAGATTTCTGAGTTCAAGGCCACAGCTGTCTATAGAGT	4680
Db	4621	GCACCTTGGAGGAGAAAGCAGGAGATTTCTGAGTTCAAGGCCACAGCTGTCTATAGAGT	4680
QY	4681	GAGTTTCCAGGACAGCCAGGGCTACACAGAGAAACCTGTTTGAAGAAAAACAGAAAAACA	4740
Db	4681	GAGTTTCCAGGACAGCCAGGGCTACACAGAGAAACCTGTTTGAAGAAAAACAGAAAAACA	4740
QY	4741	AACAAACAAACAAACAAACCAACCCAAACCCAAACCAACCTCTCATCTCTCATCTCTC	4800
Db	4741	AACAAACAAACAAACAAACCAACCCAAACCCAAACCAACCTCTCATCTCTCATCTCTC	4800
QY	4801	TAGGCTGTGTCTGTAGTGTAGATTTGGGAGTCTTAGACTTATATATTAATATAGGCC	4860

|||||  
Db 4801 TAGGCTGTCTCTAGGTAGAGTTTGGGACTTCAGACTATATATTAATGAGCC 4860  
QY 4861 TTTTATATCATCTGTGAGAGACAGAAAGTTTCAGTCTGTGGGACACAGTGGACCTGAGAA 4920  
Db 4861 TTTTATATCATCTGTGAGAGACAGAAAGTTTTCAGTCTGTGGGACACAGTGGACCTGAGAA 4920  
QY 4921 AAGTACTCTTCCAGCCAAAATTTCTGGAGGCTTCTGTGGAGAAAGTGTGCTCCGAT 4980  
Db 4921 AAGTACTCTTCCAGCCAAAATTTCTGTGGAGAGGCTTCTGTGGAGAAAGTGTGCTCCGAT 4980  
QY 4981 CAGACTACTGTCTAGAAAGCAGAAAGAGGGTTGGAGAAATGTTGCTGGACAGACAGTT 5040  
Db 4981 CAGACTACTGTCTAGAAAGCAGAAAGAGGGTTGGAGAAATGTTGCTGGACAGACAGTT 5040  
QY 5041 GGAACAGAGAGCAGAGGGGGAGGCATCCAGATTCTGAAATGTAGCTGACTTTTGT 5100  
Db 5041 GGAACAGAGAGCAGAGGGGGAGGCATCCAGATTCTGAAATGTAGCTGACTTTTGT 5100  
QY 5101 TCTGTGGGTGACAAAGTGTCCCGAGGATAGGGCTGTAGAAAGGGACCGAGGGGTGAGCC 5160  
Db 5101 TCTGTGGGTGACAAAGTGTCCCGAGGATAGGGCTGTAGAAAGGGACCGAGGGGTGAGCC 5160  
QY 5161 AATGACTTCAAGTTGAGGAGACATCCAGCCAGGGCTCTGTGGCAAGCTTAAGAAATG 5220  
Db 5161 AATGACTTCAAGTTGAGGAGACATCCAGCCAGGGCTCTGTGGCAAGCTTAAGAAATG 5220  
QY 5221 AAGGCTCTAACCCCTCCGTAAGTTTAGGGAGACAGAGAGCTGAGAGATCCCTCTA 5280  
Db 5221 AAGGCTCTTAACCCCTCCGTAAGTTTAGGGAGACAGAGAGCTGAGAGATCCCTCTA 5280  
QY 5281 GGGTGAAGAGAGATATCTGTCTGTGACAAATGGCTAGAGACAGAAAGAGTGTGACAG 5340  
Db 5281 GGGTGAAGAGAGATATCTGTCTGTGACAAATGGCTAGAGACAGAAAGAGTGTGACAG 5340  
QY 5341 TTAACCCCTCAGAACAGGACCATCCCTTGGCTTAAGAGAGCTGGGCCCTTTCTGTTT 5400  
Db 5341 TTAACCCCTCAGAACAGGACCATCCCTTGGCTTAAGAGAGCTGGGCCCTTTCTGTTT 5400  
QY 5401 AAGAACTTACTTTTCTTCAGAGAGAGGAGGAGGCTTTGGCCCTCCCTGTGGTCA 5460  
Db 5401 AAGAACTTACTTTTCTTCAGAGAGAGGAGGAGGCTTTGGCCCTCCCTGTGGTCA 5460  
QY 5461 TAAACACCCCTGTGTAACTTAATTTTACTGTGCTAGTTGCTCCAGAGACGTCA 5520  
Db 5461 TAAACACCCCTGTGTAACTTAATTTTACTGTGCTAGTTGCTCCAGAGACGTCA 5520  
QY 5521 TCTGTAGACCTGTCTCTAACTACCAAGGTATGGCCACATTCCTCACCAGAAAG 5580  
Db 5521 TCTGTAGACCTGTCTCTAACTACCAAGGTATGGCCACATTCCTCACCAGAAAG 5580  
QY 5581 TGCACAAAGAGACCTTAGGAAAGGTTACATTAACAAAGATGGCCAGAAATAAACAA 5640  
Db 5581 TGCACAAAGAGACCTTAGGAAAGGTTACATTAACAAAGATGGCCAGAAATAAACAA 5640  
QY 5641 ACTACTATCCCTTGTAAACCAATTTGCTTGTGAACAGAGAGGGGTGTGAGTGTAT 5700  
Db 5641 ACTACTATCCCTTGTAAACCAATTTGCTTGTGAACAGAGAGGGGTGTGAGTGTAT 5700  
QY 5701 GT 5760  
Db 5701 GT 5760  
QY 5761 CTGGGGGACCTTTCATGCTAAGAAATCTGATATTTGGCGCCATGCCAAAGGGGTAT 5820  
Db 5761 CTGGGGGACCTTTCATGCTAAGAAATCTGATATTTGGCGCCATGCCAAAGGGGTAT 5820  
QY 5821 TGGGAGAGAGTCAAGGCTTGTGCAAAACAGTAAGCTGCCAAGATGATGTGTGCTGAA 5880  
Db 5821 TGGGAGAGAGTCAAGGCTTGTGCAAAACAGTAAGCTGCCAAGATGATGTGTGCTGAA 5880  
QY 5881 TTACCAAGGGGAGGCTGATCAGAGTGGACAGAACATCAAGATTAAGCCACCTGTGG 5940  
|||||

Db 5881 TTACCAAGGGGAGGCTGATCAGAGTGGACAGAACATCAAGATTAAGCCACCTGTGG 5940  
QY 5941 GCTCAGAAAGAGGAGTTTACAGAGGTAAAGGCCAAAGCATTTATTTTCCAGACATGAC 6000  
Db 5941 GCTCAGAAAGAGGAGTTTACAGAGGTAAAGGCCAAAGCATTTATTTATTCAGACATGAC 6000  
QY 6001 TCAAAATCAAAAGTGCAGAGAGATTTAGCTGTGAGAGATGGGCTGTCAGTGTGGACACC 6060  
Db 6001 TCAAAATCAAAAGTGCAGAGAGATTTAGCTGTGAGAGATGGGCTGTCAGTGTGGACACC 6060  
QY 6061 TGAACCTTGACATTTATTTAGTCTAGAGCCAAAGAGAGATCACAGAGGTGATGGTCTTA 6120  
Db 6061 TGAACCTTGACATTTATTTAGTCTAGAGCCAAAGAGAGATCACAGAGGTGATGGTCTTA 6120  
QY 6121 CTGAGCTTGAGACAGGACAGTGGAGAAATGGGAGCTTCATCCTGATGAGAGGGCTGAG 6180  
Db 6121 CTGAGCTTGAGACAGGACAGTGGAGAAATGGGAGCTTCATCCTGATGAGAGGGCTGAG 6180  
QY 6181 CACCAACAGGTACAAGTGTCCCTGTGTCTCATGCCAGAGATTCCTGGCCAGTTTCAAG 6240  
Db 6181 CACCAACAGGTACAAGTGTCCCTGTGTCTCATGCCAGAGATTCCTGGCCAGTTTCAAG 6240  
QY 6241 GACTAAGGACTATCTGTGTGGAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6300  
Db 6241 GACTAAGGACTATCTGTGTGGAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6300  
QY 6301 AATTTAATCAGAAACCCCTGGGATGAGGCTGAGGACAGAGAGCTTTTAAAAAGCTC 6360  
Db 6301 AATTTAATCAGAAACCCCTGGGATGAGGCTGAGGACAGAGAGCTTTTAAAAAGCTC 6360  
QY 6361 CCAGGTATTTGATCAGAGCTGGAACAAACAGAGCTCAGGTTTCAAAACAGAAAGAGC 6420  
Db 6361 CCAGGTATTTGATCAGAGCTGGAACAAACAGAGCTCAGGTTTCAAAACAGAAAGAGC 6420  
QY 6421 AAAGCTAAGGAAAGCTTGGGATGGGAGGCTTTCAGGCCAGTATGAGGCTGTT 6480  
Db 6421 AAAGCTAAGGAAAGCTTGGGATGGGAGGCTTTCAGGCCAGTATGAGGCTGTT 6480  
QY 6481 AAGCTAAGGAAAGCTTGGGATGGGAGGCTTTCAGGCCAGTATGAGGCTGTT 6540  
Db 6481 AAGCTAAGGAAAGCTTGGGATGGGAGGCTTTCAGGCCAGTATGAGGCTGTT 6540  
QY 6541 ACACCCACCCATCCATTTATGACACCATCTTCATCCATCCATCATGATCAGCTACCCAC 6600  
Db 6541 ACACCCACCCATCCATTTATGACACCATCTTCATCCATCCATCATGATCAGCTACCCAC 6600  
QY 6601 CCAGGATCCATCAAAACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Db 6601 CCAGGATCCATCAAAACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
QY 6661 CATTTATCCACAGAACTGGTATTTGATTAAGTGGGAGATTTAATTTTATTTTGA 6720  
Db 6661 CATTTATCCACAGAACTGGTATTTGATTAAGTGGGAGATTTAATTTTATTTTGA 6720  
QY 6721 ACCTGTGTTGATTTGACTGATTTGATGATGATGATGATGATGATGATGATGATGATG 6780  
Db 6721 ACCTGTGTTGATTTGACTGATTTGATGATGATGATGATGATGATGATGATGATGATG 6780  
QY 6781 TGGCAATCGGAGAAAGTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6840  
Db 6781 TGGCAATCGGAGAAAGTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6840  
QY 6841 TTGAACCTCAAAATTTATGCGGCTGTGGCAAGTCTTTTACACCGAGCCATTTTGTCTG 6900  
Db 6841 TTGAACCTCAAAATTTATGCGGCTGTGGCAAGTCTTTTACACCGAGCCATTTTGTCTG 6900  
QY 6901 CATCATTTATTTATTAAGAAAGCATTTATGATGATGATGATGATGATGATGATGATGAT 6960  
Db 6901 CATCATTTATTTATTAAGAAAGCATTTATGATGATGATGATGATGATGATGATGATGAT 6960  
QY 6961 CCAGGATGAGCTTTTAACTCTGCTCTTCACGCTTCACCGAGTGTATGATTTACAGGT 7020  
Db 6961 CCAGGATGAGCTTTTAACTCTGCTCTTCACGCTTCACCGAGTGTATGATTTACAGGT 7020

Qy	7021	GTTAACTGGTGAATGCCCTTTAATCCAGACTCTGTGGGGGGGGGGGGAGCGGATC	7080
Db	7020	GTTCAACTGGGTGAATGCCCTTTAATCCACACACTCTGTGGGGGGGGGGGGAGCGGATC	7079
Qy	7081	CCTGAGTTGGAGGCGACTTTGGTGTCTACAGAGTTTCAGATCTCTGGGCTATACAGGAA	7140
Db	7080	CCTGAGTTGGAGGCGACTTTGGTGTCTACAGAGTTTCAGATCTCTGGGCTATACAGGAA	7139
Qy	7141	ACCTATCCCAACAAACAACAACAACAAAAATATCTGTGCATTAATCACAGAGA	7200
Db	7140	ACCTATCCCAACAAACAACAACAACAAAAATATCTGTGCATTAATCACAGAGA	7199
Qy	7201	TTAGAGGATATTACTAGGAGTAGTAGGGCGTGGAGGAGAGTCATGCTTCTTTGTATT	7260
Db	7200	TTAGAGGATATTACTAGGAGTAGTAGGGCGTGGAGGAGAGTCATGCTTCTTTGTATT	7259
Qy	7261	ATAATAGTAAGTAACTGCACAGATGCATTATCTATCTATCTATCTATCTATCTATC	7320
Db	7260	ATAATAGTAAGTAACTGCACAGATGCATTATCTATCTATCTATCTATCTATCTATC	7319
Qy	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC	7380
Db	7320	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC	7379
Qy	7381	CTGCTTTGACCTCAATAGCTCTCTATTCTGGGGCACTCTTACCCTAGAGTTGGGTT	7440
Db	7380	CTGCTTTGACCTCAATAGCTCTCTATTCTGGGGCACTCTTACCCTAGAGTTGGGTT	7439
Qy	7441	ACCAACACCCAGACATTATTTTATTTTATTTTATTTATCTATCTATCTATCTATCTATC	7500
Db	7440	ACCAACACCCAGACATTATTTTATTTTATTTTATTTATCTATCTATCTATCTATCTATC	7499
Qy	7501	GGACTCAGGGGCTGTGTCATGCTAAGCAAGCTCTGCCACAGACTGAGCTCAGCTC	7560
Db	7500	GGACTCAGGGGCTGTGTCATGCTAAGCAAGCTCTGCCACAGACTGAGCTCAGCTC	7559
Qy	7561	CCATTTTGTTCAGGTGACTGTGTGACAGTTGTCTATTTGGCAGCGCTATGAGCTCTC	7620
Db	7560	CCATTTTGTTCAGGTGACTGTGTGACAGTTGTCTATTTGGCAGCGCTATGAGCTCTC	7619
Qy	7621	CACCTCCAGTCCAGCACTTTCTGTGCAATCCAGTGGGGGGGGCAACCTGTGCTCACA	7680
Db	7620	CACCTCCAGTCCAGCACTTTCTGTGCAATCCAGTGGGGGGGGCAACCTGTGCTCACA	7679
Qy	7681	GTCGCTGTTCCTCGTCTTCAGACCTACATATTTGGCTCTCGAACAGTTCATGTAATG	7740
Db	7680	GTCGCTGTTCCTCGTCTTCAGACCTACATATTTGGCTCTCGAACAGTTCATGTAATG	7739
Qy	7741	GGATGCGTTCCTGTGTATTTCTTTTATGGCTGGGCCCTTATCTTACACAGTTGTGTG	7800
Db	7740	GGATGCGTTCCTGTGTATTTCTTTTATGGCTGGGCCCTTATCTTACACAGTTGTGTG	7799
Qy	7801	GGCAGTGTGCACGCTATCTCTATCTTATCATCATCTTATGGCTTAAATAGTGTCTT	7860
Db	7800	GGCAGTGTGCACGCTATCTCTATCTTATCATCATCTTATGGCTTAAATAGTGTCTT	7859
Qy	7861	TGTGTGATAAACCACCTTCTGTTCATTTACTGATGAATTTGTGGCCACCCACCAC	7920
Db	7860	TGTGTGATAAACCACCTTCTGTTCATTTACTGATGAATTTGTGGCCACCCACCAC	7919
Qy	7921	CCTTTTTTTTTTTATTTGACACAAGGCTTCTGTGTATCTTGTCAATCTTGGCGTCT	7980
Db	7920	CCTTTTTTTTTTTATTTGACACAAGGCTTCTGTGTATCTTGTCAATCTTGGCGTCT	7979
Qy	7981	GAGCTCACTCTGTAGACACAGGCTGTAGGCTGTCTTCCACTTTTGACACTCTGTGAAC	8040
Db	7980	GAGCTCACTCTGTAGACACAGGCTGTAGGCTGTCTTCCACTTTTGACACTCTGTGAAC	8039
Qy	8041	AGAGTACCATGAACCTTCAAGACAAATTTTCTGTGTTTGTGTTTTTACATTTGTGTG	8100
Db	8040	AGAGTACCATGAACCTTCAAGACAAATTTTCTGTGTTTGTGTTTTTACATTTGTGTG	8098

QY	8101	TGTAATGCTGTAATATATGCAATGTTTGTGCTTACAGTGTCTGCAATGTGTATGCTGTGTG	8160
Db	8099	TGTATGCTGTATATGTATGCAATGTTTGTGTCTTACAGTGTCTGCAATGTGTATGCTGTGTG	8158
QY	8161	TGGACACAGAAACAACCCGATGTGACCATTCCTCAGATATACAGCATCTTGTTAATATGTA	8220
Db	8159	TGGACACAGAAACAACCCGATGTGACCATTCCTCAGATATACAGCATCTTGTTAATATGTA	8218
QY	8221	TGTAATATGTAATGTTTATTTAGTGTGCCAAGTATGCAAGTAATTTTGTGTGAGTTTTCAC	8280
Db	8219	TGTAATATGTAATGTTTATTTAGTGTGCCAAGATGCAAGTAATTTTGTGTGAGTTTTCAC	8278
QY	8281	CTTCCACTGTGGGCTCCGCCATTAAATCAGATCTCTGGGTAGTAGAGCAATGGCTTCA	8340
Db	8279	CTTCCACTGTGGGCTCTCCGCCATTAAATCAGATCTCTGGGTAGTAGAGCAATGGCTTCA	8338
QY	8341	CTCGATGAGCCATCTCGCTCCGCCCTCTGSCCACTTCCTCTTATTTTCCAGATGGAGTA	8400
Db	8339	CTCGATGAGCCATCTCGCTCCGCCCTCTGSCCACTTCCTCTTATTTTCCAGATGGAGTA	8398
QY	8401	CGCAGCTGCATCTGGCTCTTAAGCTTACCAGTCAATCCAGATGGCTTAGCCAGGAGACTCAG	8460
Db	8399	CGCAGCTGCATCTGGCTCTTAAGCTTACCAGTCAATCCAGATGGCTTAGCCAGGAGACTCAG	8458
QY	8461	GGATATGCTGGGCTCTGCGCTCCACAGTGTGTAAATTTACAGGATATCATCTGCTGGAG	8520
Db	8459	GGATATGCTGGGCTCTGCGCTCCACAGTGTGTAAATTTACAGGATATCATCTGCTGGAG	8518
QY	8521	ATTTTAACTGAATCTCTGAGATAGAGCAGGCACTCTACCAATGGAGGTTCTTTTGT	8580
Db	8519	ATTTTAACTGAATCTCTGAGATAGAGCAGGCACTCTACCAATGGAGGTTCTTTTGT	8578
QY	8581	GTTTGGTTGGTTTCCCTCCATATAGATCAGGAGTGTCAATTAGTAGGCTGGGCTAC	8640
Db	8579	GTTTGGTTGGTTTCCCTCTGCAATATAGATCAGGAGTGTCAATTAGTAGGCTGGGCTAC	8638
QY	8641	ATAACATCTTGTCTCAAAAAGCCTATAGAGGATAGGAGGTCGAGGCTAAGAAGAGCCT	8700
Db	8639	ATAACATCTTGTCTCAAAAAGCCTATAGAGGATAGGAGGTCGAGGCTAAGAAGAGCCT	8698
QY	8701	AAGCGGCTGTGATGATGACACACAGATAGGCTGACATATATACCAAGACCTTGTTCAAA	8760
Db	8699	AAGCGGCTGTGATGACACACAGATAGGCTGACATATATACCAAGACCTTGTTCAAA	8758
QY	8761	ACATGAGGAGGAGGGATGTTTTAAGTGTGTGGCTGTGTAAACGACACTAAGAGGAGCCA	8820
Db	8759	ACATGAGGAGGAGGGATGTTTTAAGTGTGTGGCTGTGTAAACGACACTAAGAGGAGCCA	8818
QY	8821	TGTATGACATTTGACTAAGAAAGATATATCATCAAGCCGGGTGGGCAAGGATAGAGTTGG	8880
Db	8819	TGTATGACATTTGACTAAGAAAGATATATCATCAAGCCGGGTGGGCAAGGATAGAGTTGG	8878
QY	8881	ACTACAGTGTCTAAAGACCCCATAGAGAACCAAGTTCCCTTTCTCTGGGCTCAAGC	8940
Db	8879	ACTACAGTGTCTAAAGACCCCATAGAGAACCAAGTTCCCTTTCTCTGGGCTCAAGC	8938
QY	8941	CTGGCTCGACGCGCCTGCTCTCATATGCCCTTCTCCTTAGGCTGTCCACCATAG	8995
Db	8939	CTGGCTCGACGCGCCTGCTCTCATATGCCCTTCTCCTTAGGCTGTCCACCATAG	8993

RESULT 3	AA19044	standard	CDNA: 2628 BP.
ID	AA19044		
XX			
XX			
AC	AA19044:		
XX			
DT	09-APR-2002	(first entry)	
XX			
DE	Mouse Mu-1 haematopoietin receptor superfamily chain CDNA sequence.		
XX			
XX			
MM	Mouse; MU-1; haematopoietin receptor superfamily chain; thyroiditis		
KW	immune deficiency; anaemia; autoimmune disorder; multiple sclerosis		



```
XX 22-OCT-2001 (first entry)
XX
XX
DE Human polynucleotide SEQ ID NO 5076.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0596042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HSE-) HXSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang J, Mehman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPT: 2001-442253/47.
XX P-PSDB: AAM41931.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1: SEQ ID NO 5076; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 2447 BP; 733 A; 387 C; 515 G; 812 T; 0 other;
XX
XX Query Match 0.7%; Score 64; DB 22; Length 2447;
XX Best Local Similarity 100.0%; Pred. No. 8.6e-17;
XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 6
AAS13715/C
ID AAS13715 standard; DNA; 80 BP.
XX
XX AAS13715;
XX
XX 08-MAY-2002 (first entry)
XX
XX Simple sequence repeat, SSR, #12.
XX
XX Simple sequence repeat; plant; ds; SSR; ryegrass; fescue; tandem repeat;
XX cereal profiling; grass profiling; seed batch purity testing.
XX
XX Poeae.
XX
XX NZ509193-A.
XX
XX 25-MAY-2001.
XX
XX 03-JAN-2001; 2001NZ-0509193.
XX
XX 24-DEC-1999; 99AU-0004906.
XX
XX 04-MAY-2000; 2000AU-0007310.
XX
XX (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
XX (USC-) UNIV SOUTHERN CROSS.
XX (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
XX (UVAD-) UNIV ADELAIDE.
XX (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
XX
XX Forster JW, Jones ES;
XX
XX WPT: 2001-512563/56.
XX
XX New simple sequence repeats having 2 or more tandemly repeated
XX nucleotide core elements isolated from ryegrass and fescue, useful for
XX selecting of genes in grass or cereal breeding or profiling grass or
XX cereal species varieties -
XX
XX Claim 6; Page 51; 72pp; English.
XX
XX The invention relates to a substantially purified or isolated nucleic
XX acid (I) from ryegrass or fescue species including a simple sequence
XX repeat (SSR), having 2 or more tandemly repeated nucleotide core elements
XX 2-6 nucleotides in length. Also included are a nucleic acid primer
XX suitable for amplifying an SSR, identifying (M) an SSR by preparing a
XX library of ryegrass or fescue genomic DNA enriched for SSRs and
XX identifying clones in the library containing SSRs, a library of ryegrass
XX or fescue genomic DNA enriched for SSRs prepared by the M, selecting for
XX a gene in grass or cereal breeding by identifying an SSR that is closely
XX associated with the gene such that the SSR and the gene are
XX preferentially co-inherited, and selecting for the SSR in the
XX breeding, a method for DNA profiling grass or cereal species varieties by
XX assessing variation between SSR varieties and testing the purity of grass
XX or cereal seed batches by assessing variation within seed batch of an
XX SSR. The SSRs may be used in the selection of genes in grass or cereal
XX breeding, for profiling grass or cereal species varieties, for testing
XX the purity of grass or cereal seed batches, and for DNA profiling to
XX establish the distinct identity, uniformity and/or stability of a
XX cultivar. The present sequence is a ryegrass or fescue SSR.
XX
XX Sequence 80 BP; 40 A; 33 C; 0 G; 7 T; 0 other;
XX
XX Query Match 0.7%; Score 63; DB 23; Length 80;
XX Best Local Similarity 100.0%; Pred. No. 2.9e-16;
XX Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```











Db 1605 GT 1604

Search completed: January 29, 2003, 14:09:44  
Job time : 1956 secs

---



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2003, 11:46:17 : Search time 21787 Seconds

(without alignments)  
12015.410 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatcgtgtgcacccaagaca.....ctctagctcgtccacatg 8995

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8995	100.0	8995	6	AX027824	AX027824 Sequence
2	4465	49.6	192060	2	AC098570	AC098570 Mus muscu
3	96	1.1	149109	2	AC101018	AC101018 Rattus no
4	96	1.1	159492	2	AC115173	AC115173 Rattus no
5	82	0.9	222185	2	AC127697	AC127697 Mus muscu
6	82	0.9	226909	2	AC079490	AC079490 Mus muscu
7	82	0.9	239077	2	AC093351	AC093351 Mus muscu
8	80	0.9	235302	2	AC073784	AC073784 Mus muscu
9	79	0.9	202855	10	AL589737	AL589737 Mouse DNA
10	78	0.9	204068	2	AC103395	AC103395 Mus muscu
11	77	0.9	174982	2	AC110034	AC110034 Mus muscu
12	77	0.9	184854	2	AC122433	AC122433 Mus muscu
13	76	0.8	178032	2	AC087869	AC087869 Mus muscu
14	76	0.8	201602	2	AC074146	AC074146 Mus muscu
15	76	0.8	203083	2	AC069459	AC069459 Mus muscu
16	76	0.8	245958	2	AL731687	AL731687 Mus muscu
17	75	0.8	148180	2	AC113443	AC113443 Mus muscu
18	75	0.8	168889	10	AL731779	AL731779 Mouse DNA
19	75	0.8	191080	2	AC083914	AC083914 Mus muscu
20	75	0.8	200370	2	AC112940	AC112940 Mus muscu
21	75	0.8	263350	2	AC130210	AC130210 Mus muscu
22	74	0.8	178428	2	AC108601	AC108601 Rattus no
23	74	0.8	221378	2	AC122835	AC122835 Mus muscu
24	74	0.8	271723	2	AC122210	AC122210 Mus muscu
25	73	0.8	15000	10	AF420002	AF420002 Mus muscu
26	73	0.8	110000	2	AC125151_2	Continuation (3 of
27	73	0.8	187271	2	AC102745	AC102745 Mus muscu
28	73	0.8	192086	2	AC117634	AC117634 Mus muscu
29	73	0.8	205336	10	AC116576	AC116576 Mus muscu
30	73	0.8	222973	2	AC093403	AC093403 Mus muscu
31	73	0.8	238186	2	AC112676	AC112676 Mus muscu
32	71	0.8	106605	9	AL353691	AL353691 Human DNA
33	71	0.8	110716	2	AC097770	AC097770 Rattus no
34	71	0.8	153727	10	AC093449	AC093449 Mus muscu
35	71	0.8	166509	2	AC117101	AC117101 Rattus no
36	71	0.8	191895	2	AL844872	AL844872 Mus muscu
37	71	0.8	197632	10	AL672039	AL672039 Mouse DNA
38	71	0.8	214000	10	AC093317	AC093317 Mus muscu
39	70	0.8	149930	2	AC130543	AC130543 Mus muscu
40	70	0.8	150552	2	AC121952	AC121952 Mus muscu
41	70	0.8	176814	10	AL627444	AL627444 Mouse DNA
42	70	0.8	184167	2	AC118211	AC118211 Mus muscu
43	70	0.8	205615	2	AC117826	AC117826 Mus muscu
44	70	0.8	211096	2	AC127309	AC127309 Mus muscu
45	70	0.8	211348	10	AC098883	AC098883 Mus muscu

#### ALIGNMENTS

RESULT 1  
AX027824 LOCUS AX027824 8995 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent WO0034492.  
ACCESSION AX027824  
VERSION AX027824.1 GI:10188668  
KEYWORDS  
SOURCE  
ORGANISM Mus sp.  
Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 8995)  
AUTHORS Robine,S., Louvard,D., Pinto,D. and Jaisser,F.  
TITLE Regulatory sequences of the mouse villin gene - use in transgenesis  
JOURNAL Patent: WO 0034492-A 1 15-JUN-2000;

Pred. No. is the number of results predicted by chance to have a



Db 1921 AGGACTCATATGGTGGCTCACAGCCATCTGTAAATCCAGTTCCAGAGGGTTCCACACCTT 1980  
Qy 1981 CTTCTGGCCTCCACAGGCACACATATAGTACACAGACATACATGCGAGGCAAAACACC 2040  
Db 1981 CTTCTGGCCTCCACAGGCACACATATAGTACACAGACATACATGCGAGGCAAAACACC 2040  
Qy 2041 CATACACATATAATTAATTAAGAACTTAAAGGTGTCATGTGTTGTTAAACATTTGTCT 2100  
Db 2041 CATACACATATAATTAATTAAGAACTTAAAGGTGTCATGTGTTGTTAAACATTTGTCT 2100  
Qy 2101 TACACATGCTGATGTAAGACATGTACACGACACACTGAAGAGGATCTGGGGCTGGAG 2160  
Db 2101 TACACATGCTGATGTAAGACATGTACACGACACACTGAAGAGGATCTGGGGCTGGAG 2160  
Qy 2161 AGATGGCTCAGGGGTTAAGAGCAGCTAGCTCTTCCGAAGGAAGGTCCGATTTCAAT 2220  
Db 2161 AGATGGCTCAGGGGTTAAGAGCAGCTAGCTCTTCCGAAGGAAGGTCCGATTTCAAT 2220  
Qy 2221 CCTAGCACCCATAGTGGCTCACAACCATCCATTAATGAGATGACACCCCTTCTGTG 2280  
Db 2221 CCTAGCACCCATAGTGGCTCACAACCATCCATTAATGAGATGACACCCCTTCTGTG 2280  
Qy 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTAGATATACTAATTAATTAATCTTTT 2340  
Db 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTAGATATACTAATTAATTAATCTTTT 2340  
Qy 2341 TTTAAAAAATGAAGAGGATCTGACACCTGAAAAAGATTAATAGCAGTACACG 2400  
Db 2341 TTTAAAAAATGAAGAGGATCTGACACCTGAAAAAGATTAATAGCAGTACACG 2400  
Qy 2401 GGTGATTATCTATCTGAGATTTTTCCTTCCGCTTGGCTTGCACAGTGGTGAGACAG 2460  
Db 2401 GGTGATTATCTATCTGAGATTTTTCCTTCCGCTTGGCTTGCACAGTGGTGAGACAG 2460  
Qy 2461 CCCCTTTTCACTTACAGAAGCGGGCTACATTAATTTCTGACAAAAACAGCACCCTCAGT 2520  
Db 2461 CCCCTTTTCACTTACAGAAGCGGGCTACATTAATTTCTGACAAAAACAGCACCCTCAGT 2520  
Qy 2521 ATGTACTGTCTGCTGCTGATGAGCAGCGGCGGGCGGACACACACACAC 2580  
Db 2521 ATGTACTGTCTGCTGCTGATGAGCAGCGGCGGGCGGACACACACACAC 2580  
Qy 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTGGGAAAGTCA 2640  
Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTGGGAAAGTCA 2640  
Qy 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTATCTTCCCTCTAAGAGACACAGATTC 2700  
Db 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTATCTTCCCTCTAAGAGACACAGATTC 2700  
Qy 2701 AAGGTGGCAGAAGATCTACAGGGGGCAGAGGAGGGGGAAGAGGCAATGATTC 2760  
Db 2701 AAGGTGGCAGAAGATCTACAGGGGGCAGAGGAGGGGGAAGAGGCAATGATTC 2760  
Qy 2761 AGAGACCTACAGCAGAGGCGCAGCAAGAGATCCAGGTCCAGGGCAGGGAGGAGTGA 2820  
Db 2761 AGAGACCTACAGCAGAGGCGCAGCAAGAGATCCAGGTCCAGGGCAGGGAGGAGTGA 2820  
Qy 2821 CCCTTTCTCCGAGAGAGGCGGCGCAGAGAGGTTCAAGGCAAGGTTATGSCA 2880  
Db 2821 CCCTTTCTCCGAGAGAGGCGGCGCAGAGAGGTTCAAGGCAAGGTTATGSCA 2880  
Qy 2881 GGTCAATAAAGTGGAGGTGGTGCATCTCAGAAAGAGAGAAAGGAAAGGCCCTTGT 2940  
Db 2881 GGTCAATAAAGTGGAGGTGGTGCATCTCAGAAAGAGAGAAAGGAAAGGCCCTTGT 2940  
Qy 2941 GCCCAGTGAAGAGGTCATGCTGAGTAGAGATCTGCAGGGGTGCCAGAGGCCCA 3000  
Db 2941 GCCCAGTGAAGAGGTCATGCTGAGTAGAGATCTGCAGGGGTGCCAGAGGCCCA 3000  
Qy 3001 CTCTCTGTCCTCAAGGAAACCCCAAGTGTGAATCTTGCCCTGGCTGAGTTCCAGCTA 3060  
Db 3001 CTCTCTGTCCTCAAGGAAACCCCAAGTGTGAATCTTGCCCTGGCTGAGTTCCAGCTA 3060

Qy 3061 CAAGACCCCAAGAGTGTCTACTCTCATCCATCCAGTCCAGTGGCCCTCGGCCGCCACACCCA 3120  
Db 3061 CAAGACCCCAAGAGTGTCTACTCTCATCCATCCAGTCCAGTGGCCCTCGGCCGCCACACCCA 3120  
Qy 3121 CCCCAGCTCCCTGCGCAGCTTCTCTAGGGCTGAGAGGTGGCGAGCCCTGGTGGGGTTGC 3180  
Db 3121 CCCCAGCTCCCTGCGCAGCTTCTCTAGGGCTGAGAGGTGGCGAGCCCTGGTGGGGTTGC 3180  
Qy 3181 CTACTGCAAGGTAGAGCCAGGTCTAGCCCGGAAGTGCACACCCATCCCTGAAGCTGCAGA 3240  
Db 3181 CTACTGCAAGGTAGAGCCAGGTCTAGCCCGGAAGTGCACACCCATCCCTGAAGCTGCAGA 3240  
Qy 3241 GCCAAGGGGGGGCACAACGGCAGCTCAGGCTGTCAGAGCTGTTCCTGGCTCTAGTTCC 3300  
Db 3241 GCCAAGGGGGGGCACAACGGCAGCTCAGGCTGTCAGAGCTGTTCCTGGCTCTAGTTCC 3300  
Qy 3301 AGGACCTGGGCGACCTACTTCCACCCCCCATCCATCTCTCTGGGGCCATCTTCC 3360  
Db 3301 AGGACCTGGGCGACCTACTTCCACCCCCCATCCATCTCTCTGGGGCCATCTTCC 3360  
Qy 3361 CTTATATGCTGAAGAAATTCTGGGGGGGGGGGTGTGTGAGACAAAGTCTTGC 3420  
Db 3361 CTTATATGCTGAAGAAATTCTGGGGGGGGGGGTGTGTGAGACAAAGTCTTGC 3420  
Qy 3421 GTCTCTGCAAGCCAGCTTGGCACAACCTTCTAGATCTCCAGTGTGGTGGCTCCCTTC 3480  
Db 3421 GTCTCTGCAAGCCAGCTTGGCACAACCTTCTAGATCTCCAGTGTGGTGGCTCCCTTC 3480  
Qy 3481 CAGACAGTGAAGCAATTGGGTGGGACACATGTATGCCACAGTGGTTGAGAGGAGAC 3540  
Db 3481 CAGACAGTGAAGCAATTGGGTGGGACACATGTATGCCACAGTGGTTGAGAGGAGAC 3540  
Qy 3541 GGTCTTCTCTCTCTGCGAGCTGTGCTTCTGTAGACACTTGGTATTAAGTTGGGG 3600  
Db 3541 GGTCTTCTCTCTCTGCGAGCTGTGCTTCTGTAGACACTTGGTATTAAGTTGGGG 3600  
Qy 3601 TGAAGTAAAGTGTCTGAAGCTGAAAGCTGAAGAGCAAGAGCAGCAGGCTGTGGGCT 3660  
Db 3601 TGAAGTAAAGTGTCTGAAGCTGAAAGCTGAAAGAGCAAGAGCAGCAGGCTGTGGGCT 3660  
Qy 3661 TCAATGAAGAAATTCACAGACCCCTTCTCTGTAAGTCACTTCCCTCATCTGTAG 3720  
Db 3661 TCAATGAAGAAATTCACAGACCCCTTCTCTGTAAGTCACTTCCCTCATCTGTAG 3720  
Qy 3721 ATTCCCTGGAGCAAGGTGCTCTGGGACTCAGATTTCTACATTTAAATCAGGACAGT 3780  
Db 3721 ATTCCCTGGAGCAAGGTGCTCTGGGACTCAGATTTCTACATTTAAATCAGGACAGT 3780  
Qy 3781 CCTGAGACTTGGACCTCGGCGCTGATTTACTACTCTCTGCTGGCTGCTCATTTCTGTG 3840  
Db 3781 CCTGAGACTTGGACCTCGGCGCTGATTTACTACTCTCTGCTGGCTGCTCATTTCTGTG 3840  
Qy 3841 TCATGTCTTACACATCTGAATAGTTCTTGTGTACCAATCCCTGACACTCTGGGA 3900  
Db 3841 TCATGTCTTACACATCTGAATAGTTCTTGTGTACCAATCCCTGACACTCTGGGA 3900  
Qy 3901 GGTGATCTCTTGGCACAATGTATCTGGGATGTAACTGTGACGCCACAGAGAGAGGGG 3960  
Db 3901 GGTGATCTCTTGGCACAATGTATCTGGGATGTAACTGTGACGCCACAGAGAGAGGGG 3960  
Qy 3961 AGAGTGAAGAGTGTCTGCTAGGCTCTATTAGGCTGAGACATACGCCCTTCTCTAGAAAT 4020  
Db 3961 AGAGTGAAGAGTGTCTGCTAGGCTCTATTAGGCTGAGACATACGCCCTTCTCTAGAAAT 4020  
Qy 4021 GGCCCTCCATTTTTCGTTACCATGATCTATTTTATACAGAGTGGGAGTGAAGGCA 4080  
Db 4021 GGCCCTCCATTTTTCGTTACCATGATCTATTTTATACAGAGTGGGAGTGAAGGCA 4080  
Qy 4081 AACCTGCCAGAAATTTGGGACTCACTCAGACCAGGTTATCTGCTCAGAAATCCCTTG 4140  
Db 4081 AACCTGCCAGAAATTTGGGACTCACTCAGACCAGGTTATCTGCTCAGAAATCCCTTG 4140





Db 6301 AATTAAATCAGAAACCCCTGGGATGCGAGGCTTCGACGACGAGGAGCTTTTAAAAAGCTC 6360  
OY 6361 CCAGGTATTTCTGATGAGCAGCTGGAACAACACAGCTACAGGTTCAACAGAGAAGGCG 6420  
Db 6361 CCAGGTATTTCTGATGAGCAGCTGGAACAACACAGCTACAGGTTCAACAGAGAAGGCG 6420  
OY 6421 AAAGCTAGGAAAAGCTTGGGATG6GGAGGCTTTCAGGCGCAGTAGATGAGAGCTGTT 6480  
Db 6421 AAAGCTAGGAAAAGCTTGGGATG6GGAGGCTTTCAGGCGCAGTAGATGAGAGCTGTT 6480  
OY 6481 AGCAGTGTGGGAGCTTCTCTGCGCTGTCAATATAGCTATCCATCCAGCTATCCATCAT 6540  
Db 6481 AGCAGTGTGGGAGCTTCTCTGCGCTGTCAATATAGCTATCCATCCAGCTATCCATCAT 6540  
OY 6541 ACACCCACCCATCCATTTATGACCCATCCTTCATCCATCCATCCATCCATCCATCCATCC 6600  
Db 6541 ACACCCACCCATCCATTTATGACCCATCCTTCATCCATCCATCCATCCATCCATCCATCC 6600  
OY 6601 CCACGATCCATCCAAACCTTCCCTTTCCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 6660  
Db 6601 CCACGATCCATCCAAACCTTCCCTTTCCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 6660  
OY 6661 CATTTATCCACAGAGACTGATGATGATTAATGTGGAGATTTAATTAATTTTGA 6720  
Db 6661 CATTTATCCACAGAGACTGATGATGATTAATGTGGAGATTTAATTAATTTTGA 6720  
OY 6721 AGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
Db 6721 AGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
OY 6781 TGGCAATCGGAGAAAGTTTGGGTTGTTTCTTCTTCCACGCTGGGTTCTGGGA 6840  
Db 6781 TGGCAATCGGAGAAAGTTTGGGTTGTTTCTTCTTCCACGCTGGGTTCTGGGA 6840  
OY 6841 TTGACTCAATATATGCGGCTGTGCAAGTCTTTACCACGACCAATTTTGTGACA 6900  
Db 6841 TTGAACTCAATATATGCGGCTGTGCAAGTCTTTACCACGACCAATTTTGTGACA 6900  
OY 6901 CATCATTTATTTAGAAAGATTTATGATGATGATGATGATGATGATGATGATGATGATG 6960  
Db 6901 CATCATTTATTTAGAAAGATTTATGATGATGATGATGATGATGATGATGATGATGATG 6960  
OY 6961 CCACGATGACCTTTTAACTCTGCTCTTCCAGGCTCCACCGAGCTAGGTTTACAGGT 7020  
Db 6961 CCACGATGACCTTTTAACTCTGCTCTTCCAGGCTCCACCGAGCTAGGTTTACAGGT 7020  
OY 7021 GTTCAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7080  
Db 7021 GTTCAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7080  
OY 7081 CCGAGTTGGAGGCGAGTTTGTCTACAGATTTCAAGATACCTGGGCGTATCCAGGAA 7140  
Db 7081 CCGAGTTGGAGGCGAGTTTGTCTACAGATTTCAAGATACCTGGGCGTATCCAGGAA 7140  
OY 7141 ACCCTATCCCAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7200  
Db 7141 ACCCTATCCCAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7200  
OY 7201 TTAGAGGATTTAGTGGGATGATGAGGCTGTGGAGGAGATGATGATGATGATGATG 7260  
Db 7201 TTAGAGGATTTAGTGGGATGATGAGGCTGTGGAGGAGATGATGATGATGATGATG 7260  
OY 7261 ATTAATGTAAGTACACAGAATGATTAATCTATCTATCTATCTATCTATCTATCTATC 7320  
Db 7261 ATTAATGTAAGTACACAGAATGATTAATCTATCTATCTATCTATCTATCTATCTATC 7320  
OY 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380  
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380  
OY 7381 CTGCTTTGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7440  
Db 7381 CTGCTTTGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7440

OY 7441 ACCAACCCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 7500  
Db 7441 ACCAACCCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 7500  
OY 7501 GGACTCAGGGCTTGTGATGCTTAAGCAAGCTCTGCCACAGAGCTGCAAGCTCAGTCC 7560  
Db 7501 GGACTCAGGGCTTGTGATGCTTAAGCAAGCTCTGCCACAGAGCTGCAAGCTCAGTCC 7560  
OY 7561 CCATTTTGTTCAGGTACCTGTGACAGTTGTATATTGGCAGCGTATGTAGCTCTCTC 7620  
Db 7561 CCATTTTGTTCAGGTACCTGTGACAGTTGTATATTGGCAGCGTATGTAGCTCTCTC 7620  
OY 7621 CACCTCCAGTTCCAGACACTTTCGTGATCCAGAGGGGGGCACTGTGCTCACA 7680  
Db 7621 CACCTCCAGTTCCAGACACTTTCGTGATCCAGAGGGGGGCACTGTGCTCACA 7680  
OY 7681 GTGCCCTGTCCCTGTCTGACACTACATATTGGCTGTGTAAGAGTTTCATGTAATG 7740  
Db 7681 GTGCCCTGTCCCTGTCTGACACTACATATTGGCTGTGTAAGAGTTTCATGTAATG 7740  
OY 7741 GATGCTTCTCTGTATTTCTTTATGCTGGCCCTTTATCTTACACAGTTTGTCTG 7800  
Db 7741 GATGCTTCTCTGTATTTCTTTATGCTGGCCCTTTATCTTACACAGTTTGTCTG 7800  
OY 7801 GGCCATGTGACACTGCTATACCTATCTATCATCAATCAATGAGCTTAATAGTCCCT 7860  
Db 7801 GGCCATGTGACACTGCTATACCTATCTATCATCAATCAATGAGCTTAATAGTCCCT 7860  
OY 7861 TGTGTGATTAACCACTTCTGTTCATTTACAGATGAATTTGTGGCCCAACCCCAAC 7920  
Db 7861 TGTGTGATTAACCACTTCTGTTCATTTACAGATGAATTTGTGGCCCAACCCCAAC 7920  
OY 7921 CCTTTTTTTTTTATTTAGACAGAGCTTCTGTGTATCTTGCATCTTGCTCTCT 7980  
Db 7921 CCTTTTTTTTTTATTTAGACAGAGCTTCTGTGTATCTTGCATCTTGCTCTCTCT 7980  
OY 7981 GAGCTACTGTGACAGCAGGCTGTGAGGCTGCTCCACTTTTACACTCTCTGTGAC 8040  
Db 7981 GAGCTACTGTGACAGCAGGCTGTGAGGCTGCTCCACTTTTACACTCTCTGTGAC 8040  
OY 8041 AGAGTACCATGAACTTCAAGACAATTTTGTGTTTGTGTTTATTTACATTTGTG 8100  
Db 8041 AGAGTACCATGAACTTCAAGACAATTTTGTGTTTGTGTTTATTTACATTTGTG 8100  
OY 8101 TGTATGCTGTATATGTGATGTTTGTCTTACAGTCTCAGATGTGTGTGTG 8160  
Db 8101 TGTATGCTGTGTATATGTGATGTTTGTCTTACAGTGTGTGTGTGTGTGTG 8160  
OY 8161 TGGGACAGAGAACAAACCGATGTGCAATCTCAGATPACTAGCATCTTTATATATGTA 8220  
Db 8161 TGGGACAGAGAACAAACCGATGTGCAATCTCAGATPACTAGCATCTTTATATATGTA 8220  
OY 8221 TGTATATGATGTTTATTTAGTGTGCCCAAGTATCAGGATTTTGTGAGATTTTCA 8280  
Db 8221 TGTATATGATGTTTATTTAGTGTGCCCAAGTATCAGGATTTTGTGAGATTTTCA 8280  
OY 8281 CTTCCTTGTGGCTCTCCGATTTAACTACGCTCTGGGCTAGTGAGCAATGCTTTA 8340  
Db 8281 CTTCCTTGTGGCTCTCCGATTTAACTACGCTCTGGGCTAGTGAGCAATGCTTTA 8340  
OY 8341 CTGAGTGAAGCATCTGAGGCGCTGCTGACCTCTCTCTCTTTTCCAGATGGAGTA 8400  
Db 8341 CTGAGTGAAGCATCTGAGGCGCTGCTGACCTCTCTCTCTTTTCCAGATGGAGTA 8400  
OY 8401 CGCAGTCACTGGCTTAAAGCTCACCAAGTATCATCAGATGCTAGCCAGAGACTCAG 8460  
Db 8401 CGCAGTCACTGGCTTAAAGCTCACCAAGTATCATCAGATGCTAGCCAGAGACTCAG 8460  
OY 8461 GGTATATGCTGGCTCTGCTCCACAGTGTGAATTTACAGGCTATACATCACTGTGTGA 8520  
Db 8461 GGTATATGCTGGCTCTGCTCCACAGTGTGAATTTACAGGCTATACATCACTGTGTGA 8520





OY 1861 GAGATGGCTACGCTTCAGAGAGCACTTGGCTCTTGACAGAGGACCTAGATTTCAGTTCCC 1920  
|||||  
Db 42021 GAGATGGCTACGCTTCAGAGAGCACTTGGCTCTTGACAGAGGACCTAGATTTCAGTTCCC 42080  
OY 1921 AGCACTCATATGGTGGCTCACAGCCATCTGTAAATCCAGTTCGAGGGTTCACACCCCT 1980  
|||||  
Db 42081 AGCACTCATATGGTGGCTCACAGCCATCTGTAAATCCAGTTCGAGGGTTCACACCCCT 42140  
OY 1981 CTTTCGGCTCCACAGGACACATACATATGTATACAGACATATGACGGCAAAACACC 2040  
|||||  
Db 42141 CTTTCGGCTCCACAGGACACATATGTATACAGACATATGACGGCAAAACACC 42200  
OY 2041 CATACACATATAATATAAGAACTTAAAGGTGATGTGTTGGTAAACATTTGGCT 2100  
|||||  
Db 42201 CATACACATATAATATAAGAACTTAAAGGTGATGTGTTGGTAAACATTTGGCT 42260  
OY 2101 TACACATGCTGATTTGAAGACATGTATACAGCACACTGAAGGGATCTGGGGCTGGAG 2160  
|||||  
Db 42261 TACACATGCTGATTTGAAGACATGTATACAGCACACTGAAGGGATCTGGGGCTGGAG 42320  
OY 2161 AGATGGCTAGGGGTTAAAGACATGATCTCTTCCGAAGGAGAGTCCGTGAATTCGAAT 2220  
|||||  
Db 42321 AGATGGCTAGGGGTTAAAGACATGATCTCTTCCGAAGGAGAGTCCGTGAATTCGAAT 42380  
OY 2221 CCTAGCAACCATGCTGGCTCACAAACCATCATATGAGATCTGACACCCCTCTGGT 2280  
|||||  
Db 42381 CCTAGCAACCATGCTGGCTCACAAACCATCATATGAGATCTGACACCCCTCTGGT 42440  
OY 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTATGATATACTAATAATAATCTTTT 2340  
|||||  
Db 42441 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTATGATATACTAATAATAATCTTTT 42500  
OY 2341 TTTAAAAAATGAAGGGATCTGAGACACCTCAAAAGGATATAGAGACAGTACACAG 2400  
|||||  
Db 42501 TTTAAAAAATGAAGGGATCTGAGACACCTCAAAAGGATATAGAGACAGTACACAG 42560  
OY 2401 GGTGATTATCTATCTCGAGATTTTCTTCCGCTTGGCTGCAACCTGGGTGAGACAGAG 2460  
|||||  
Db 42561 GGTGATTATCTATCTCGAGATTTTCTTCCGCTTGGCTGCAACCTGGGTGAGACAGAG 42620  
OY 2461 CCCCTTTTCAATTCACAGAAAGGGGTCTACATTAATTTTGAAACAAAACAGCACTCTCAGT 2520  
|||||  
Db 42621 CCCCTTTTCAATTCACAGAAAGGGGTCTACATTAATTTTGAAACAAAACAGCACTCTCAGT 42680  
OY 2521 ATGTTACTGTCTGCTGCTACTATGAGCAAGGGCAGCGCGGCGC--CACACACACAC 2578  
|||||  
Db 42681 ATGTTACTGTCTGCTGCTACTATGAGCAAGGGCAGCGCGGCGC--CACACACACAC 42740  
OY 2579 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGT 2638  
|||||  
Db 42741 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGT 42800  
OY 2639 CAAGAAGAGGCTGCCCTCAAAACAGATCTTTCATCTTCCCTCTTAAAGAGCACAGATTT 2698  
|||||  
Db 42801 CAAGAAGAGGCTGCCCTCAAAACAGATCTTTCATCTTCCCTCTTAAAGAGCACAGATTT 42860  
OY 2699 CCAAGGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGGGAAGGAGGCCATGGTTT 2758  
|||||  
Db 42861 CCAAGGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGGGAAGGAGGCCATGGTTT 42920  
OY 2759 CCAAGAGACCTTACAGCAGAGGGGCAAGAGCAGATCCCAAGTCCACAGGGAGGGAGTGA 2818  
|||||  
Db 42921 CCAAGAGACCTTACAGCAGAGGGGCAAGAGCAGATCCCAAGTCCACAGGGAGGGAGTGA 42980  
OY 2819 GGGCTTTGTTCCGAGAGAGAGGAGGCGCAGACAGGGTTCAAGAGCACAGGTTTATGG 2878  
|||||  
Db 42981 GGGCTTTGTTCCGAGAGAGAGGAGGCGCAGACAGGGTTCAAGAGCACAGGTTTATGG 43040  
OY 2879 CAGCTCATATAAGTGGAGTGTGGCTCACTCAGAAAGAGAGGAAGGAAGGAAAGGCCCTT 2938  
|||||  
Db 43041 CAGCTCATATAAGTGGAGTGTGGCTCACTCAGAAAGAGAGGAAGGAAGGAAAGGCCCTT 43100

OY 2939 GTGCCACTGAGCAGAGGGTCTATGCTAGTAGAGAGATCTGCAGGGGTCCAGAGACCC 2998  
|||||  
Db 43101 GTGCCACTGAGCAGAGGGTCTATGCTAGTAGAGAGATCTGCAGGGGTCCAGAGACCC 43160  
OY 2999 ACCTGTCTGTCCCAAGAGAACCCCAAGTGTGAACCTTGCGCTTGGGTGCTGAGTTCCAG 3058  
|||||  
Db 43161 ACCTGTCTGTCCCAAGAGAGAACCCCAAGTGTGAACCTTGCGCTTGGGTGCTGAGTTCCAG 43220  
OY 3059 TACAGAGCCCGAGAGTCTCTACTCCATCCCAACATGATCCCGCCCGCCGACACCC 3118  
|||||  
Db 43221 TACAGAGCCCGAGAGTCTCTACTCCATCCCAACATGATCCCGCCCGCCGACACCC 43280  
OY 3119 CACCCCGCATCCCGTGCCTTCTCTAGGGCTGAGGGTGGCCAGCCCTGCTGGGGGTT 3178  
|||||  
Db 43281 CACCCCGCATCCCGTGCCTTCTCTAGGGCTGAGGGTGGCCAGCCCTGCTGGGGGTT 43340  
OY 3179 GCTTACTCTGAGGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCCCTGAAGCTGCA 3238  
|||||  
Db 43341 GCTTACTCTGAGGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCCCTGAAGCTGCA 43400  
OY 3239 GAGCCAAAGGGGGGGCAGACAGGCGCTCAGGCTGATCAGGCTGTGGGCTCTAGGTTTC 3298  
|||||  
Db 43401 GAGCCAAAGGGGGGGCAGACAGGCGCTCAGGCTGATCAGGCTGTGGGCTCTAGGTTTC 43460  
OY 3299 CCAGGAGACCTGGGCACTACTTCCCAACCCCAATCCATCTCTCTGGGGCCCTATCTT 3358  
|||||  
Db 43461 CCAGGAGACCTGGGCACTACTTCCCAACCCCAATCCATCTCTCTGGGGCCCTATCTT 43520  
OY 3359 CCCTTATATGTGAAGAAATTCCTGGGGGGGGGGGGTGTGTGAGCAAAAGTCTGTT 3418  
|||||  
Db 43521 CCCTTATATGTGAAGAAATTCCT--GGGGGGGGGGTGTGTGAGCAAAAGTCTGTT 43578  
OY 3419 GGGTCTCTGAGCAGCAGCTTGGCACAACCTTCTTAAGATCTCCAGGTGGTGGCTGCTCT 3478  
|||||  
Db 43579 GGGTCTCTGAGCAGCAGCTTGGCACAACCTTCTTAAGATCTCCAGGTGGTGGCTGCTCT 43638  
OY 3479 TCCAGACAGTAAAGCAATTTGGGTGGGAGACACATGGTGAACAGAGTGGTTGGAGGGAG 3538  
|||||  
Db 43639 TCCAGACAGTAAAGCAATTTGGGTGGGAGACACATGGTGAACAGAGTGGTTGGAGGGAG 43698  
OY 3539 AAGGTCTCTGCTTCTCTCTGCGACGCTGTCTTCTGTAGCACTTGGTATTAAGTTGGG 3598  
|||||  
Db 43699 AAGGTCTCTGCTTCTCTCTGCGACGCTGTCTTCTGTAGCACTTGGTATTAAGTTGGG 43758  
OY 3599 GGTGAGGTAAAGGCTCTGAACCTTGAAGAAGCAAGAGGAGGAGGCTGCTTGGGC 3658  
|||||  
Db 43759 GGTGAGGTAAAGGCTCTGAACCTTGAAGAAGCAAGAGGAGGAGGCTGCTTGGGC 43818  
OY 3659 CTTCAATGAAGGAAGTTTCACAGACCCCTTCTCTGAAGTCACTTGGCTTCACTGTGT 3718  
|||||  
Db 43819 CTTCAATGAAGGAAGTTTCACAGACCCCTTCTCTGAAGTCACTTGGCTTCACTGTGT 43878  
OY 3719 ACATTTCCCTGGGACCAAGTGGCTCTGGGACTCAGATTTTCACAATTAATTCAGGACA 3778  
|||||  
Db 43879 ACATTTCCCTGGGACCAAGTGGCTCTGGGACTCAGATTTTCACAATTAATTCAGGACA 43938  
OY 3779 GTCTGAGACTTGGACTCCGCTGCTATTTACTACTTCTCTGAGCTCTCATATTCTGT 3838  
|||||  
Db 43939 GTCTGAGACTTGGACTCCGCTGCTATTTACTACTTCTCTGAGCTCTCATATTCTGT 43998  
OY 3839 GTTCATGTCTTACATCTGAATAGTTTCTTGTGTCACTTCCCTGAGACTCTCTGG 3898  
|||||  
Db 43999 GTTCATGTCTTACATCTGAATAGTTTCTTGTGTCACTTCCCTGAGACTCTCTGG 44058  
OY 3899 GAGGTCTGTATCTTGGACATGTATCTTGGATGTAGCTGACGCCACAGAGAGAGGG 3958  
|||||  
Db 44059 GAGGTCTGTATCTTGGACATGTATCTTGGATGTAGCTGACGCCACAGAGAGAGGG 44118  
OY 3959 GAGAGGTTCAGAGCTGTGCTTACAGGCCCTAATAGGCTGTGACATACCCCTTTCATAGA 4018  
|||||  
Db 44119 GAGAGGTTCAGAGCTGTGCTTACAGGCCCTAATAGGCTGTGACATACCCCTTTCATAGA 44178  
OY 4019 ATGGCCCTCTCATTTTTCGTTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGC 4078



```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 46829: contig of 46829 bp in length
* 46830 46929: gap of 100 bp
* 46830 49621: contig of 2692 bp in length
* 49622 49721: gap of 100 bp
* 49722 70009: contig of 20288 bp in length
* 70010 70109: gap of 100 bp
* 70110 77196: contig of 7087 bp in length
* 77197 77296: gap of 100 bp
* 77297 99453: contig of 22157 bp in length
* 99454 99553: gap of 100 bp
* 99554 102879: contig of 3326 bp in length
* 102880 102979: gap of 100 bp
* 102980 120352: contig of 17373 bp in length
* 120353 120452: gap of 100 bp
* 120453 131346: contig of 11094 bp in length
* 131547 131646: gap of 100 bp
* 131647 149109: contig of 17463 bp in length.
*
* Location/Qualifiers
*   1..149109
*     /organism="Rattus norvegicus"
*     /db_xref="taxon:10116"
*     /clone_1fb="RPC1-32 Male Rat BAC"
*     /clone_1lb="RPC1-32 Male Rat BAC"
*
BASE COUNT      39036 a 34316 c 35185 g 39719 t      853 others
ORIGIN
Query Match      1.1%; Score 96; DB 2; Length 149109;
Best Local Similarity 100.0%; Pred. No. 9e-41;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5953 GAGTTTCAAGAGAGTAAAGCCAGCATTTATTCACAAAGATGCTCAAAATCAAG 6012
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 102361 GAGTTTCAAGAGAGTAAAGCCAGCATTTATTCACAAAGATGCTCAAAATCAAG 102420

QY 6013 TGCAGAGAGAGTAAAGCCAGCATTTATTCACAAAGATGCTCAAAATCAAG 6048
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 102421 TGCAGAGAGAGTAAAGCCAGCATTTATTCACAAAGATGCTCAAAATCAAG 102456

RESULT 4
AC115173/1 159492 bp DNA linear HTG 13-JUL-2002
LOCUS      Rattus norvegicus clone CH230-286017, *** SEQUENCING IN PROGRESS
DEFINITION *** 46 unordered pieces.
AC115173
VERSION    AC115173.2 GI:21738030
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 159492)
            Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alstrooms,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Bunyah,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Dreper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

```

```

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabris,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Hayak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Horn,J.F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssohn,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Medora,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkenwo,S., Ogun,W., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svalok,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.P., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 159492)
Worley,K.C.
Direct Submission
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 159492)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:19482237.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GODN
Center clone name: CH230-286017
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124398 bases at least Q40
Consensus quality: 128059 bases at least Q30
Consensus quality: 131015 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1014: contig of 1014 bp in length
* 1015 1114: gap of unknown length
* 1115 2395: contig of 1281 bp in length
* 2396 2495: gap of unknown length

```

```

* 2496 3600: contig of 1105 bp in length
* 3601 3700: gap of unknown length
* 3701 5021: contig of 1321 bp in length
* 5022 5121: gap of unknown length
* 5122 6727: contig of 1606 bp in length
* 6728 6827: gap of unknown length
* 6828 8608: contig of 1781 bp in length
* 8609 8709 10003: contig of 1295 bp in length
* 10004 10103: gap of unknown length
* 10104 11211: contig of 1108 bp in length
* 11212 11311: gap of unknown length
* 11312 12518: contig of 1207 bp in length
* 12519 12618: gap of unknown length
* 12619 13843: contig of 1225 bp in length
* 13844 15404: gap of unknown length
* 15405 15504: gap of unknown length
* 15505 17647: contig of 2143 bp in length
* 17648 17747: gap of unknown length
* 17748 19706: contig of 1959 bp in length
* 19707 19806: gap of unknown length
* 19807 21576: contig of 1770 bp in length
* 21577 21676: gap of unknown length
* 21677 22806: contig of 1230 bp in length
* 22907 23006: gap of unknown length
* 23007 25086: contig of 2080 bp in length
* 25087 25186: gap of unknown length
* 25187 26832: contig of 1646 bp in length
* 26833 26932: gap of unknown length
* 26933 28710: contig of 1778 bp in length
* 28711 28810: gap of unknown length
* 28811 30606: contig of 1796 bp in length
* 30607 30706: gap of unknown length
* 30707 33113: contig of 2407 bp in length
* 33114 33213: gap of unknown length
* 33214 35361: contig of 2148 bp in length
* 35362 35461: gap of unknown length
* 35462 37520: contig of 2059 bp in length
* 37521 37620: gap of unknown length
* 37621 40471: contig of 2851 bp in length
* 40472 40571: gap of unknown length
* 40572 42584: contig of 2013 bp in length
* 42585 42684: gap of unknown length
* 42685 45547: contig of 2863 bp in length
* 45548 45647: gap of unknown length
* 45648 47838: contig of 2191 bp in length
* 47839 47938: gap of unknown length
* 47939 51229: contig of 3291 bp in length
* 51230 51329: gap of unknown length
* 51330 54005: contig of 2676 bp in length
* 54006 54105: gap of unknown length
* 54106 56683: contig of 2578 bp in length
* 56684 56783: gap of unknown length
* 56784 59361: contig of 2578 bp in length
* 59362 59461: gap of unknown length
* 59462 61588: contig of 2127 bp in length
* 61589 61688: gap of unknown length
* 61689 65712: contig of 4024 bp in length
* 65713 65812: gap of unknown length
* 65813 69339: contig of 3527 bp in length
* 69340 69439: gap of unknown length
* 69440 72570: contig of 3131 bp in length
* 72571 72670: gap of unknown length
* 72671 76898: contig of 4228 bp in length
* 76899 76998: gap of unknown length
* 76999 80760: contig of 3762 bp in length
* 80761 80860: gap of unknown length
* 80861 86113: contig of 5453 bp in length
* 86114 86413: gap of unknown length
* 86414 92231: contig of 5818 bp in length
* 92232 92331: gap of unknown length
* 92332 99692: contig of 7361 bp in length

```

```

* 99693 99792: gap of unknown length
* 99793 106178: contig of 6386 bp in length
* 106179 106278: gap of unknown length
* 106279 114304: contig of 8026 bp in length
* 114305 114404: gap of unknown length
* 114405 122112: contig of 7708 bp in length
* 122113 122212: gap of unknown length
* 122213 129642: contig of 7430 bp in length
* 129643 129742: gap of unknown length
* 129743 138684: contig of 8942 bp in length
* 138685 138784: gap of unknown length
* 138785 149662: contig of 10878 bp in length
* 149663 149762: gap of unknown length
* 149763 159492: contig of 9730 bp in length.
FEATURES
source 1..159492
location/Qualifiers
  /organism="Rattus norvegicus"
  /db_xref="taxon:10116"
  /clone="CH330-286017"
BASE COUNT 39083 a 36447 c 36055 g 40464 t 7443 others
ORIGIN
Query Match 1.1%; Score 96; DB 2: Length 159492;
Best Local Similarity 100.0%; Pred. No. 9e-41;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5953 GAGTTTACAGAGAGGTTAAAGCCCATTTATTCACAGACATGACCAAAATCAAG 6012
DB 27228 GAGTTTACAGAGAGGTTAAAGCCCATTTATTCACAGACATGACCAAAATCAAG 27169
QY 6013 TCACAAGAGAGATTAAGTCGAGAGATGGCGCTGCA 6048
DB 27168 TCGAAGGAGAGATTAAGTCGAGAGATGGCGCTGCA 27133
RESULT 5
AC127697 222185 bp DNA linear HTG 22-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP24-83110, WORKING DRAFT
DEFINITION AC127697.2 GI:21913472
AC127697.2
SEQUENCE 13 unordered pieces.
AC127697.2 GI:21913472
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 222185)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 222185)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 222185)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 21, 2002 this sequence version replaced gi:21903681.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0083110
-----
----- Summary Statistics -----
Sequencing vector: M13; 0%

```



Sequencing vector: plasmid; 100%  
Chemistry: Dye-Primer ET; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 217367 bases at least Q40  
Consensus quality: 218557 bases at least Q30  
Consensus quality: 219259 bases at least Q20  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved. 1042: contig of 1042 bp in length  
\* 1  
\* 1043 1142: gap of unknown length  
\* 1143 3141: contig of 1999 bp in length  
\* 3142 3241: gap of unknown length  
\* 3242 7211: contig of 3970 bp in length  
\* 7212 10408: contig of 3097 bp in length  
\* 10409 10508: gap of unknown length  
\* 10509 17015: contig of 6507 bp in length  
\* 17016 17115: gap of unknown length  
\* 17116 27558: contig of 10443 bp in length  
\* 27559 27658: gap of unknown length  
\* 27659 41472: contig of 13614 bp in length  
\* 41473 41572: gap of unknown length  
\* 41573 58617: contig of 17045 bp in length  
\* 58618 83266: contig of 24549 bp in length  
\* 83267 83366: gap of unknown length  
\* 83367 130000: contig of 46634 bp in length  
\* 130001 130100: gap of unknown length  
\* 130101 172736: contig of 42636 bp in length  
\* 172737 172836: gap of unknown length  
\* 172837 221107: contig of 48271 bp in length  
\* 221108 221208: gap of unknown length  
\* 221208 222185: contig of 978 bp in length.  
Location/Qualifiers  
1. 222185  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP24-83110"  
1. 1042  
/note="assembly\_name:Contig25"  
1143. 3141  
/note="assembly\_name:Contig26"  
3242. 7211  
/note="assembly\_name:Contig27"  
7312. 10408  
/note="assembly\_name:Contig28"  
10509. 17015  
/note="assembly\_name:Contig29"  
17116. 27558  
/note="assembly\_name:Contig30"  
27659. 41472  
/note="assembly\_name:Contig31"  
41573. 58617  
/note="assembly\_name:Contig32"  
58718. 83266  
/note="assembly\_name:Contig33"  
83367. 130000  
/note="assembly\_name:Contig34"  
130101. 172736  
/note="assembly\_name:Contig35"  
172837. 222107  
/note="assembly\_name:Contig36"  
221208. 222185  
/note="assembly\_name:Contig37"

BASE COUNT 67728 a 41550 c 41568 g 70139 t 1200 others  
ORIGIN  
Query Match 0.9%; Score 82; DB 2; Length 222185;  
Best Local Similarity 100.0%; Pred. No. 4.3e-33;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 4593 GCAGTGGTGACACAGCCCTTATCCACACTGTGGAGACAGAGCAGATTTCG 4652  
|||||  
DB 26396 GCAGTGGTGACACAGCCCTTATCCACACTGTGGAGACAGAGCAGATTTCG 26455  
|||||  
QY 4653 AGTTCAGGCCAGCCCTGTCTA 4674  
|||||  
DB 26456 AGTTCAGGCCAGCCCTGTCTA 26477  
|||||  
RESULT 6  
AC079490/c 226909 bp DNA linear HTG 02-SEP-2000  
LOCUS AC079490  
DEFINITION Mus musculus clone RP23-190A2, WORKING DRAFT SEQUENCE, 53 unordered  
pieces.  
AC079490  
VERSION AC079490.1 GI:9964855  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 226909)  
DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 226909)  
REFERENCE DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 1809834  
Center clone name: RPCI-23\_190A2  
-----  
Summary Statistics  
Consensus quality: 180161 bases at least Q40  
Consensus quality: 198601 bases at least Q30  
Consensus quality: 203733 bases at least Q20  
Estimated insert size: 209000; agarose-fp estimation  
Estimated insert size: 221709; sum-of-contigs estimation  
Quality coverage: 6.2 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.84 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 53 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved. 1  
\* 1121 1120: contig of 1120 bp in length  
\* 1221 1220: gap of unknown length  
\* 1221 2574: contig of 1334 bp in length  
\* 2575 2674: gap of unknown length  
\* 2675 4062: contig of 1388 bp in length  
\* 4063 4163: gap of unknown length  
\* 4163 5667: contig of 1504 bp in length  
\* 5667 5767: gap of unknown length  
\* 5767 6895: contig of 1129 bp in length  
\* 6895 6995: gap of unknown length  
\* 6996 8025: contig of 1030 bp in length



8026	8125	gap of unknown length
8126	9388	contig of 1263 bp in length
9389	9488	gap of unknown length
9489	10710	contig of 1222 bp in length
10711	10810	gap of unknown length
10811	11811	contig of 1001 bp in length
11812	11811	gap of unknown length
11912	12865	contig of 1054 bp in length
12866	12965	gap of unknown length
13066	14068	contig of 1203 bp in length
14269	14368	gap of unknown length
14369	15046	contig of 1478 bp in length
15847	15846	gap of unknown length
15947	17219	contig of 1273 bp in length
17220	17219	gap of unknown length
17320	18672	contig of 1353 bp in length
18673	18772	gap of unknown length
18773	19999	contig of 1227 bp in length
20000	20099	gap of unknown length
20100	21459	contig of 1360 bp in length
21460	21559	gap of unknown length
21560	22822	contig of 1063 bp in length
22823	22722	gap of unknown length
22723	23726	contig of 1004 bp in length
23727	23826	gap of unknown length
23827	25149	contig of 1323 bp in length
25150	25249	gap of unknown length
25250	26639	contig of 1590 bp in length
26840	26939	gap of unknown length
26940	28630	contig of 1691 bp in length
28631	28730	gap of unknown length
28731	30265	contig of 1535 bp in length
30266	30365	gap of unknown length
30366	31703	contig of 1338 bp in length
31704	31803	gap of unknown length
31804	32951	contig of 1148 bp in length
32952	33052	gap of unknown length
33052	33529	contig of 2278 bp in length
33530	35429	gap of unknown length
35430	38496	contig of 3067 bp in length
38497	38596	gap of unknown length
38597	40521	contig of 1925 bp in length
40522	40621	gap of unknown length
40622	42212	contig of 1591 bp in length
42213	42212	gap of unknown length
42213	44441	contig of 2129 bp in length
44442	44541	gap of unknown length
44542	46397	contig of 1856 bp in length
46398	46397	gap of unknown length
46498	48685	contig of 2188 bp in length
48686	48785	gap of unknown length
48786	50811	contig of 2026 bp in length
50812	50911	gap of unknown length
50912	52840	contig of 1729 bp in length
52841	52740	gap of unknown length
52741	53525	contig of 1212 bp in length
53526	54052	gap of unknown length
54053	54053	gap of unknown length
54054	58018	contig of 3966 bp in length
58019	58118	gap of unknown length
58119	61336	contig of 3318 bp in length
61340	61536	gap of unknown length
61537	64032	contig of 2496 bp in length
64033	64132	gap of unknown length
64133	66752	contig of 2620 bp in length
66753	66852	gap of unknown length
66853	70097	contig of 3245 bp in length
70098	70197	gap of unknown length
70198	73503	contig of 3306 bp in length
73504	73503	gap of unknown length
73504	80346	contig of 6743 bp in length
80347	80446	gap of unknown length
80447	86466	contig of 6020 bp in length
86467	86566	gap of unknown length

FEATURES	SOURCE	Location/Qualifiers
BASE COUNT	52802 a 57780 c 58798 g 52284 t 5245 others	
ORIGIN	1..226909 /organism="Mus musculus" /db_xref="taxon:10090" /clone="RP23-190A2" /clone_lib="RP23-190A2"	
Query Match	0.9%; Score 82; DB 2; Length 226909;	
Best Local Similarity	100.0%; Pred. No. 4.3e-33;	
Matches	82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	4593 GCAGTGTGGTGACACAGCCCTTATATCCAGACACTTGGAGGAGCAGAAGCAGGACGATTCTG	4652
Db	35815 GCAGTGTGGTGACACAGCCCTTATATCCAGACACTTGGAGGAGCAGAAGCAGGACGATTCTG	35756
QY	4653 AGTTCAGGCGCCAGCCGGCTCA	4674
Db	35755 AGTTCAGGCGCCAGCCGGCTCA	35734
RESULT 7	AC093351/c	
LOCUS	AC093351	
DEFINITION	Mus musculus clone RP23-16L19, WORKING DRAFT SEQUENCE, 12 unordered pieces.	
ACCESSION	AC093351	
VERSION	AC093351.1 GI:15213896	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL.	
SOURCE	Mus musculus.	
ORGANISM	Mus musculus.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.	
TITLE	Mus musculus, clone RP23-16L19	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 239077)	
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,	
REFERENCE	Anderson, S., Barra, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,	
AUTHORS	Brown, A., Canarata, J., Campolano, A., Chang, J., Chazaro, B.,	
REFERENCE	Chapel, T., Colangelo, M., Collins, S., Collymore, A., Cook, A.,	
AUTHORS	Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,	
REFERENCE	Ferreira, P., FitzHugh, W., Gage, K., Galagan, J., Gardyna, S.,	
AUTHORS	Glade, G., Goyette, M., Graham, L., Grand-Pierre, N.,	
REFERENCE	Hagos, B., Hefford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,	
AUTHORS	Jones, C., Kanat, A., Karatas, A., Kells, C., Laroque, K.,	
REFERENCE	Lamares, R., Landers, T., Lehotzky, D., Levine, R., Liu, G.,	
AUTHORS	Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,	
REFERENCE	McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,	
AUTHORS	Meneus, L., Mihova, T., Mlenka, Y., Murphy, T., Naylor, J., Nguyen, C.,	

TITLE  
JOURNAL  
COMMENT

Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunhahang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schnupp, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (20-AUG-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L13561

Center clone name: 16\_L\_19

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 234078 bases at least Q40

Consensus quality: 237194 bases at least Q20

Insert size: 222000; agarose-fp

Insert size: 237977; sum-of-ctrls

Quality coverage: 11.5 in Q20 bases; agarose-fp  
Quality coverage: 10.7 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently  
consists of 12 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 5895: contig of 5895 bp in length  
5896 5995: gap of 100 bp  
5996 6105: contig of 110 bp in length  
6106 6205: gap of 100 bp  
6206 6873: contig of 668 bp in length  
6874 6973: gap of 100 bp  
6974 7992: contig of 1019 bp in length  
7993 8092: gap of 100 bp  
8093 10126: contig of 2034 bp in length  
10127 10226: gap of 100 bp  
10227 16035: contig of 5809 bp in length  
16036 16135: gap of 100 bp  
16136 30334: contig of 14199 bp in length  
30335 30434: gap of 100 bp  
30435 59946: contig of 29512 bp in length  
59947 60046: gap of 100 bp  
60047 79068: contig of 19022 bp in length  
79069 79168: gap of 100 bp  
79169 122677: contig of 43509 bp in length  
122678 122777: gap of 100 bp  
122778 176372: contig of 53595 bp in length  
176373 176473: gap of 100 bp  
176473 239077: contig of 62605 bp in length.

FEATURES  
source

1. 239077

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="RP23-16L19"

/clone\_lib="RPCT-23 Female Mouse BAC"

misc\_feature

1. 5895

/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
misc\_feature  
5996..6105  
/note="assembly\_fragment"  
misc\_feature  
6206..6873  
/note="assembly\_fragment"  
misc\_feature  
6974..7992  
/note="assembly\_fragment"  
misc\_feature  
8093..10126  
/note="assembly\_fragment"  
misc\_feature  
10227..16035  
/note="assembly\_fragment"  
misc\_feature  
16136..30334  
/note="assembly\_fragment"  
misc\_feature  
30435..59946  
/note="assembly\_fragment"  
misc\_feature  
60047..79068  
/note="assembly\_fragment"  
misc\_feature  
79169..122677  
/note="assembly\_fragment"  
misc\_feature  
122778..176372  
/note="assembly\_fragment"  
misc\_feature  
176473..239077  
/note="assembly\_fragment"

BASE COUNT 65130 a 55831 c 54944 g 62063 t 1109 others  
ORIGIN

Query Match 0.9%; Score 82; DB 2; Length 239077;  
Best Local Similarity 100.0%; Pred. No. 4.3e-33;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4638 GCAGCAGATTTCGAGTTCAGAGCCAGCCTGCTATAGAGTCCAGACAGCA 4697  
Db 42272 GCAGCAGATTTCGAGTTCAGAGCCAGCCTGCTATAGAGTCCAGACAGCA 42213  
OY 4698 GGGCTACACAGAGAAACCTGT 4719  
Db 42212 GGGCTACACAGAGAAACCTGT 42191

RESULT 8  
AC073784 235302 bp DNA linear HTG 29-JUN-2000  
LOCUS Mus musculus clone RP23-412G23, WORKING DRAFT SEQUENCE, 46  
DEFINITION Unordered pieces.  
ACCESSION AC073784.1 GI:8810401  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Mus musculus.  
SOURCE Mus musculus.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 235302)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 235302)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

Project Information  
Center Project Name: 1895247  
Center clone name: RPCT-23\_412G23  
-----  
Summary Statistics  
Consensus quality: 197724 bases at least Q40

Consensus quality: 214621 bases at least Q30  
 Consensus quality: 218454 bases at least Q20  
 Estimated insert size: 203000; agarose-1p estimation  
 Estimated insert size: 230802; sum-of-ctdigs estimation  
 Quality coverage: 7.31 in Q20 bases; agarose-1p estimation  
 Quality coverage: 6.43 in Q20 bases; sum-of-ctdigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 46 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1442: contig of 1442 bp in length  
 \* 1443 1542: gap of unknown length  
 \* 1543 2717: contig of 1175 bp in length  
 \* 2718 2817: gap of unknown length  
 \* 2818 3834: contig of 1017 bp in length  
 \* 3835 3934: gap of unknown length  
 \* 3935 4966: contig of 1032 bp in length  
 \* 4967 5066: gap of unknown length  
 \* 5067 6129: contig of 1063 bp in length  
 \* 6130 6229: gap of unknown length  
 \* 6230 7407: contig of 1178 bp in length  
 \* 7408 7507: gap of unknown length  
 \* 7508 8550: contig of 1043 bp in length  
 \* 8551 8650: gap of unknown length  
 \* 8651 9654: contig of 1004 bp in length  
 \* 9655 9754: gap of unknown length  
 \* 9755 11221: contig of 1467 bp in length  
 \* 11222 11321: gap of unknown length  
 \* 11322 12433: contig of 1112 bp in length  
 \* 12434 12533: gap of unknown length  
 \* 12534 13942: contig of 1409 bp in length  
 \* 13943 14042: gap of unknown length  
 \* 14043 15118: contig of 1076 bp in length  
 \* 15119 15218: gap of unknown length  
 \* 15219 17182: contig of 1964 bp in length  
 \* 17183 17282: gap of unknown length  
 \* 17283 18452: contig of 1170 bp in length  
 \* 18453 18552: gap of unknown length  
 \* 18553 20400: contig of 1848 bp in length  
 \* 20401 20500: gap of unknown length  
 \* 20501 21738: contig of 1238 bp in length  
 \* 21739 21838: gap of unknown length  
 \* 21839 23128: contig of 1290 bp in length  
 \* 23129 23228: gap of unknown length  
 \* 23229 24987: contig of 1759 bp in length  
 \* 24988 25087: gap of unknown length  
 \* 25088 26474: contig of 1387 bp in length  
 \* 26475 26574: gap of unknown length  
 \* 26575 28321: contig of 1747 bp in length  
 \* 28322 28421: gap of unknown length  
 \* 28421 30965: contig of 2544 bp in length  
 \* 30966 31065: gap of unknown length  
 \* 31066 34079: contig of 3014 bp in length  
 \* 34080 34180: gap of unknown length  
 \* 34180 36552: contig of 2373 bp in length  
 \* 36553 36652: gap of unknown length  
 \* 36653 38883: contig of 2231 bp in length  
 \* 38884 41860: gap of unknown length  
 \* 41861 41960: gap of unknown length  
 \* 41961 45013: contig of 3053 bp in length  
 \* 45014 45113: gap of unknown length  
 \* 45114 47059: contig of 1946 bp in length  
 \* 47060 47159: gap of unknown length  
 \* 47160 50539: contig of 3380 bp in length  
 \* 50540 52880: gap of unknown length  
 \* 52881 52980: contig of 2241 bp in length  
 \* 52981 55866: gap of unknown length  
 \* 55866: contig of 2886 bp in length

\* 55867 55966: gap of unknown length  
 \* 55967 59063: contig of 3097 bp in length  
 \* 59064 59163: gap of unknown length  
 \* 59164 63021: contig of 3888 bp in length  
 \* 63022 63121: gap of unknown length  
 \* 63122 69865: contig of 6744 bp in length  
 \* 69866 69965: gap of unknown length  
 \* 69966 78312: contig of 8347 bp in length  
 \* 78313 78412: gap of unknown length  
 \* 78413 85315: contig of 6903 bp in length  
 \* 85316 85415: gap of unknown length  
 \* 85416 93535: contig of 8120 bp in length  
 \* 93536 93635: gap of unknown length  
 \* 93636 104777: contig of 11142 bp in length  
 \* 104778 104877: gap of unknown length  
 \* 104878 116166: contig of 11289 bp in length  
 \* 116167 116266: gap of unknown length  
 \* 116267 128380: contig of 12114 bp in length  
 \* 128381 128480: gap of unknown length  
 \* 128481 138205: contig of 9725 bp in length  
 \* 138206 138305: gap of unknown length  
 \* 138306 152655: contig of 14260 bp in length  
 \* 152656 152665: gap of unknown length  
 \* 152666 165170: contig of 12505 bp in length  
 \* 165171 165270: gap of unknown length  
 \* 165271 177903: contig of 12633 bp in length  
 \* 177904 178003: gap of unknown length  
 \* 178003 190290: contig of 12287 bp in length  
 \* 190291 190390: gap of unknown length  
 \* 190391 211048: contig of 20658 bp in length  
 \* 211049 211148: gap of unknown length  
 \* 211149 235302: contig of 24154 bp in length.

FEATURES  
 source  
 1..235302  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="Rp23-412G23"  
 /clone\_lib="RPC1 mouse BAC library 23"  
 BASE COUNT 57053 a 57886 c 58312 g 57537 t 4514 others  
 ORIGIN

Query Match 0.9%; Score 80; DB 2; Length 235302;  
 Best Local Similarity 100.0%; Pred.No. 5.4e-32;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2202 GAAGTCCCTGAGTCAATCCCTAGCAACCATGCTGCTACACCATCCATATGAGA 2261  
 |||||||  
 DB 119422 GAAGTCCCTGAGTCAATCCCTAGCAACCATGCTGCTACACCATCCATATGAGA 119481  
 |||||||

OY 2262 TCTGACACCCCTCTTGTG 2281  
 |||||||

Db 119482 TCTGACACCCCTCTTGTG 119501

RESULT 9  
 AL589737/c  
 LOCUS 202855 bp DNA linear ROD 29-JUN-2002  
 DEFINITION Mouse DNA sequence from clone Rp23-45114 on chromosome 13, complete  
 sequence.  
 ACCESSION AL589737  
 VERSION AL589737.11 GI:17976581  
 KEYWORDS HTG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 Bates/K.  
 REFERENCE Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk  
 On Dec 21, 2001 this sequence version replaced gi:17939723.  
 COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at

<http://www.sanger.ac.uk/projects/C-elegans/wormpep> RP23-43L14 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

-----Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: <http://mrcseq.har.mrc.ac.uk>  
Contact: [mouseq@har.mrc.ac.uk](mailto:mouseq@har.mrc.ac.uk)

FEATURES  
Source Location/Qualifiers

1..202853  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/clone="RP23-43L14"  
/clone\_lib="RPI-23"  
BASE COUNT 58329 a 42347 c 42534 g 59645 t  
ORIGIN

#### Query Match

Best local similarity 100.0%; Pred. No. 1.9e-31;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4596 GTGCTGGACACGCTTTAATCCACACTGGGAGGAGAGACAGAGATTCTGAGT 4655  
|||||  
DB 40449 GTGCTGGACACGCTTTAATCCACACTGGGAGGAGAGACAGAGATTCTGAGT 40390  
QY 4656 TCAGGCGCACGCTGCTCA 4674  
|||||  
DB 40389 TCAGGCGCACGCTGCTCA 40371

#### RESULT 10

AC103395 204068 bp DNA linear HTG 21-AUG-2002  
LOCUS Mus musculus clone RP24-387L1, WORKING DRAFT SEQUENCE, 46 unordered  
DEFINITION pieces.  
AC103395  
VERSION AC103395.2 GI:22381756  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS 1 (bases 1 to 204068)  
TITLE Mus musculus, clone RP24-387L1  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 204068)  
TITLE Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cook, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

#### COMMENT

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,  
Lamzates, R., Landers, T., Lenoczky, J., Levine, R., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Struass, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggilo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

#### Direct Submission

Submitted (24-Nov-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 204068)

Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Chopel, Y., Collymore, A.,  
Cook, A., Cook, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,  
McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mienga, V.,  
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,  
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

#### Direct Submission

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced g1:17063270.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

#### Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

#### Project Information

Center project name: L19498  
Center clone name: 387\_L1  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.9607.1  
Consensus quality: 187322 bases at least Q40  
Consensus quality: 195292 bases at least Q30  
Consensus quality: 197972 bases at least Q20  
Insert size: 176000; agarose-fp  
Insert size: 199568; sum-of-contigs  
Quality coverage: 6.3 in Q20 bases; agarose-fp  
Quality coverage: 5.6 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 46 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will





```
misc_feature /note="assembly_fragment"
10507. .12527 /note="assembly_fragment"
misc_feature 12628. .14448 /note="assembly_fragment"
misc_feature 14549. .16709 /note="assembly_fragment"
misc_feature 16810. .18297 /note="assembly_fragment"
misc_feature 18398. .20768 /note="assembly_fragment"
misc_feature 20869. .23477 /note="assembly_fragment"
misc_feature 23578. .26149 /note="assembly_fragment"
misc_feature 26250. .30689 /note="assembly_fragment"
misc_feature 30790. .37563 /note="assembly_fragment"
misc_feature 37664. .47762 /note="assembly_fragment"
misc_feature 47863. .55531 /note="assembly_fragment"
misc_feature 55632. .64314 /note="assembly_fragment"
misc_feature 64415. .75211 /note="assembly_fragment"
misc_feature 75312. .87080 /note="assembly_fragment"
misc_feature 87181. .124016 /note="assembly_fragment"
```

```
Query Match 0.9% Score 77: DB 2: Length 174982:
Best Local Similarity 100.0%: Pred. NO. 2.4e-30:
Matches 77: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
```

```
QY 4638 GCAGCAGATTCTGAGTCAAGCCAGCCTGCTATAGAGTGCAGACAGCA 4697
DB 148571 GCAGCAGATTCTGAGTCAAGCCAGCCTGCTATAGAGTGCAGACAGCA 148512
QY 4698 GGGCTACACAGCAAC 4714
DB 148511 GGGCTACACAGCAAC 148495
```

```
RESULT 12
AC122433 184854 bp DNA linear HTG 06-JUL-2002
DEFINITION Mus musculus chromosome UNK clone RP24-194G3, WORKING DRAFT
ACCESSION AC122433
VERSION AC122433.2 GI:21700727
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 184854)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 184854)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
3 (bases 1 to 184854)
REFERENCE McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 6, 2002 this sequence version replaced gi:21105893.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BB0194G03
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 177563 bases at least Q40
Consensus quality: 178348 bases at least Q30
Consensus quality: 179250 bases at least Q20
Insert size: 221000; agarose-fp
Insert size: 184885; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 10.13 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1821: contig of 1821 bp in length
1822 1921: gap of unknown length
1922 5618: contig of 3697 bp in length
5619 5718: gap of unknown length
5719 11721: contig of 6003 bp in length
11722 11821: gap of unknown length
11822 18606: contig of 6785 bp in length
18607 18706: gap of unknown length
18707 23368: contig of 4662 bp in length
23369 23468: gap of unknown length
23469 31153: contig of 7685 bp in length
31154 40066: contig of 8813 bp in length
40067 40166: gap of unknown length
40167 53402: contig of 13236 bp in length
53403 53502: gap of unknown length
53503 69231: contig of 15729 bp in length
69232 69331: gap of unknown length
69332 87512: contig of 18181 bp in length
87513 87612: gap of unknown length
87613 117654: contig of 30042 bp in length
117655 117755: gap of unknown length
117756 184854: contig of 67100 bp in length.
-----
FEATURES
source
1. 184854
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-194G3"
1. 1821
/note="assembly_name:Contig10"
1922. 5618
/note="assembly_name:Contig11"
5719. 11721
/note="assembly_name:Contig12"
11822. 18606
/note="assembly_name:Contig13"
18707. 23368
/note="assembly_name:Contig14"
23469. 31153
/note="assembly_name:Contig15"
31254. 40066
/note="assembly_name:Contig16"
```

```
misc_feature      40167..53402
                  /note="assembly_name:Contig17"
misc_feature      53503..69231
                  /note="assembly_name:Contig18"
misc_feature      69332..87512
                  /note="assembly_name:Contig19"
misc_feature      87613..117654
                  /note="assembly_name:Contig20"
misc_feature      117755..184854
                  /note="assembly_name:Contig21"
BASE COUNT      49138 a 41968 c 41001 g 51557 t 1190 others
ORIGIN
Query Match      0.9%; Score 77; DB 2; Length 184854;
Best Local Similarity 100.0%; Pred. No. 2.4e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4643 CAGATTTCGAGTTCAGGCCACCCGCTGTATAGAGTTCAGTTCCAGACGACGCGCT 4702
      |||||||
DB 119686 CAGATTTCGAGTTCAGGCCACCCGCTGTATAGAGTTCAGTTCCAGACGACGCGCT 119745
      |||||||
QY 4703 ACACAGAGAACCCTGT 4719
      |||||||
DB 119746 ACACAGAGAACCCTGT 119762

RESULT 13
AC087869      178032 bp      DNA      linear      HTG 24-FEB-2001
LOCUS      Mus musculus clone RP23-297D13, *** SEQUENCING IN PROGRESS ***, 36
DEFINITION      Mus musculus clone RP23-297D13, *** SEQUENCING IN PROGRESS ***, 36
ACCESSION      AC087869
VERSION      AC087869.1 GI:12658016
KEYWORDS      HTG; HTGS; PHASE1.
SOURCE      Mus musculus.
ORGANISM      Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 178032)
McCombie,W.R., Baker,J.P., Bahret,A., Bai,H., Balija,V.,
Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,
Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., VIL,M.D.
and Zutterman,T.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 178032)
McCombie,W.R.
Direct Submission
Submitted (03-FEB-2001) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
-----
Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/geneseq
Contact: mcombie@cshl.org
-----
Project Information
Center project name: RP23-297D13
Center clone name: RP23-297D13
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 15715: contig of 15715 bp in length
* 15716 15757: gap of unknown length
* 15758 29656: contig of 13899 bp in length

FEATURES
SOURCE
1. 178032
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="RP23-297D13"
```



```
BASE COUNT      49436 a 38986 c 38427 g 49678 t 1505 others
ORIGIN

Query Match      0.8%; Score 76; DB 2; Length 178032;
Best Local Similarity 100.0%; Pred. No. 8.5e-30;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4638 GCAGGCAGATTTCGATTCAGAGCCAGCCTGCTATAGAGTGAAGTTCAGAGACGCCA 4697
      |||||||
Db 86489 GCAGGCAGATTTCGATTCAGAGCCAGCCTGCTATAGAGTGAAGTTCAGAGACGCCA 86548

Qy 4698 GCGCTACACAGAGAAA 4713
      |||||||
Db 86549 GCGCTACACAGAGAAA 86564

RESULT 14
AC074146      201602 bp      DNA      linear      HTG_15-JUL-2000
LOCUS      Mus musculus clone RP23-107G3, WORKING DRAFT SEQUENCE, 38 unordered
DEFINITION      pieces.
ACCESSION      AC074146
VERSION      AC074146.1 GI:9211256
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
ORGANISM      Mus musculus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 201602)
AUTHORS      DOE Joint Genome Institute.
TITLE      Unpublished
JOURNAL
REFERENCE      2 (bases 1 to 201602)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
      -----Genome Center
      Center: Joint Genome Institute
      Center Code: JGI
      Web site: http://www.jgi.doe.gov
      -----
      Project Information
      Center Project Name: 0
      Center clone name: RPCI-23_107G3
      -----
      Summary Statistics
      Consensus quality: 134464 bases at least Q40
      Consensus quality: 159739 bases at least Q30
      Consensus quality: 170106 bases at least Q20
      Estimated insert size: 181300; agarose-fp estimation
      Estimated insert size: 197802; sum-of-contigs estimation
      Quality coverage: 3.41 in Q20 bases; agarose-fp estimation
      Quality coverage: 3.12 in Q20 bases; sum-of-contigs estimation.
      * NOTE: This is a 'working draft' sequence. It currently
      * consists of 38 contigs. The true order of the pieces
      * is not known and their order in this sequence record is
      * arbitrary. Gaps between the contigs are represented as
      * runs of N, but the exact sizes of the gaps are unknown.
      * This record will be updated with the finished sequence
      * as soon as it is available and the accession number will
      * be preserved.
      1
      1675: contig of 1675 bp in length
      1676      1775: gap of unknown length
      1776      3291: contig of 1516 bp in length
      3292      3391: gap of unknown length
      3392      4432: contig of 1041 bp in length
      4433      4532: gap of unknown length
      4533      6053: contig of 1521 bp in length
      6054      6153: gap of unknown length
      6154      7161: contig of 1008 bp in length
      7162      7261: gap of unknown length
      7262      8417: contig of 1156 bp in length
```

```
* 8418      8517: gap of unknown length
* 8518      9800: contig of 1283 bp in length
* 9801      9900: gap of unknown length
* 9901      10926: contig of 1026 bp in length
* 10927      11026: gap of unknown length
* 11027      12062: contig of 1036 bp in length
* 12063      12162: gap of unknown length
* 12163      14110: contig of 1948 bp in length
* 14111      14210: gap of unknown length
* 14211      15534: contig of 1324 bp in length
* 15535      15634: gap of unknown length
* 15635      17316: contig of 1682 bp in length
* 17317      17416: gap of unknown length
* 17417      20982: contig of 3566 bp in length
* 20983      21082: gap of unknown length
* 21083      24936: contig of 3854 bp in length
* 24937      25036: gap of unknown length
* 25037      28124: contig of 3088 bp in length
* 28125      28424: gap of unknown length
* 28425      30948: contig of 2724 bp in length
* 30949      31048: gap of unknown length
* 31049      34011: contig of 2963 bp in length
* 34012      34111: gap of unknown length
* 34112      37382: contig of 3271 bp in length
* 37383      37482: gap of unknown length
* 37483      41136: contig of 3654 bp in length
* 41137      41236: gap of unknown length
* 41237      45754: contig of 4518 bp in length
* 45755      45854: gap of unknown length
* 45855      48474: contig of 2620 bp in length
* 48475      48574: gap of unknown length
* 48575      52059: contig of 3485 bp in length
* 52060      52159: gap of unknown length
* 52160      55993: contig of 3834 bp in length
* 55994      56093: gap of unknown length
* 56093      59136: contig of 3043 bp in length
* 59137      59236: gap of unknown length
* 59237      63015: contig of 3779 bp in length
* 63016      63115: gap of unknown length
* 63116      66817: contig of 3802 bp in length
* 66817      67017: gap of unknown length
* 67018      72225: contig of 5208 bp in length
* 72225      72325: gap of unknown length
* 72326      80793: contig of 8468 bp in length
* 80794      80893: gap of unknown length
* 80894      88129: contig of 7236 bp in length
* 88129      88229: gap of unknown length
* 88230      95766: contig of 7537 bp in length
* 95767      95866: gap of unknown length
* 95867      104692: contig of 8826 bp in length
* 104693      104792: gap of unknown length
* 104793      113453: contig of 8661 bp in length
* 113454      113553: gap of unknown length
* 113554      121609: contig of 8056 bp in length
* 121610      121709: gap of unknown length
* 121710      132117: contig of 10408 bp in length
* 132118      132217: gap of unknown length
* 132218      143836: contig of 11619 bp in length
* 143837      143936: gap of unknown length
* 143937      153823: contig of 9887 bp in length
* 153824      153923: gap of unknown length
* 153924      174899: contig of 20976 bp in length
* 174900      174999: gap of unknown length
* 175000      201602: contig of 26603 bp in length.

FEATURES
      source
      1..201602
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="RP23-107G3"
      /clone_lib="RPCI mouse BAC library 23"
BASE COUNT      52434 a 47973 c 47287 g 49607 t 4301 others
ORIGIN
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:31 : Search time 29.4042 Seconds  
(without alignments)  
12668.132 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23

Sequence: 1 gagtgcgtatgttgagagagacct 23

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estcnu:\*  
5: em\_estcov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query	Length	DB	ID	Description
	1	23	100.0	175	10	BB583035	BB583035
	2	23	100.0	225	10	BB583295	BB583295
	3	23	100.0	244	10	BB583549	BB583549
	4	23	100.0	235	10	BB582128	BB582128
	5	23	100.0	284	10	BB846624	BB846624
	6	23	100.0	354	10	BB871941	BB871941

C	7	23	100.0	414	10	BB848113	BB848113
C	8	23	100.0	524	10	AM259173	AM259173
C	9	23	100.0	637	10	BB661515	BB661515
C	10	23	100.0	674	13	B1143380	B1143380
C	11	23	100.0	752	9	A1663456	A1663456
C	12	23	100.0	764	12	BF785145	BF785145
C	13	23	100.0	856	13	B1100945	B1100945
C	14	23	100.0	879	13	B1100201	B1100201
C	15	23	100.0	897	12	BF781298	BF781298
C	16	23	100.0	916	14	BQ956792	BQ956792
C	17	23	100.0	936	14	BQ935016	BQ935016
C	18	23	100.0	956	12	BF687321	BF687321
C	19	23	100.0	997	10	BB609451	BB609451
C	20	21.4	93.0	109	9	AU076562	AU076562
C	21	21.4	93.0	300	9	AU098819	AU098819
C	22	21.4	93.0	481	9	AA070552	AA070552
C	23	21.4	93.0	710	13	B1252311	B1252311
C	24	21.4	93.0	723	13	B1761873	B1761873
C	25	21.4	93.0	796	13	B116224	B116224
C	26	21.4	93.0	875	14	BQ652678	BQ652678
C	27	21.4	93.0	903	14	BQ650039	BQ650039
C	28	21.4	93.0	935	14	BQ651090	BQ651090
C	29	21.4	93.0	952	14	BQ647976	BQ647976
C	30	21.4	93.0	969	14	BQ650225	BQ650225
C	31	21.4	93.0	1014	14	BM923688	BM923688
C	32	21.4	93.0	1206	14	BM923496	BM923496
C	33	21	91.3	413	10	BB845511	BB845511
C	34	19.8	86.1	230	10	AW336826	AW336826
C	35	19.8	86.1	441	10	BB847149	BB847149
C	36	19.8	86.1	578	10	AV607319	AV607319
C	37	18.8	81.7	726	17	AZ448995	AZ448995
C	38	18.4	80.0	443	12	BF516764	BF516764
C	39	18.4	80.0	499	12	BG275365	BG275365
C	40	18.4	80.0	581	17	AZ710407	AZ710407
C	41	18.2	79.1	397	17	AQ116749	AQ116749
C	42	18.2	79.1	407	10	AV732812	AV732812
C	43	18.2	79.1	470	10	AM823714	AM823714
C	44	18.2	79.1	486	14	BQ034151	BQ034151
C	45	18.2	79.1	493	14	BQ048109	BQ048109

#### ALIGNMENTS

RESULT 1  
BB583035/LOCUS  
DEFINITION BB583035 RIKEN full-length enriched, adult male cecum Mus musculus  
CDNA clone 9130203C06 5', mRNA sequence.  
ACCESSION BB583035  
VERSION BB583035.1 GI:11479579  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE  
AUTHORS Alizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.,  
Imotani,K., Ishii,Y., Itoh,M., Iwama,M., Kawai,J., Kojima,Y., Kono,  
H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,  
Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,  
Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shingawa,A.,  
Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,  
T., Toyota,T., Watanishi,A., Yamamura,T., Yasunishi,A., Yoshida,K.,  
Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Alizawa,K. et al. 2000)  
COMMENT  
JOURNAL Unpublished (2000)  
CONTACT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gs.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Thermostabilization and thermocatalysis of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)  
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

#### FEATURES

##### source

Location/Qualifiers  
 1. 175  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_1ib="9130203C06"  
 /clone\_1ib="RIKEN full-length enriched, adult male cecum"  
 /sex="male"  
 /tissue\_type="cecum"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGACTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTTCAGATTAAATTAATTCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 49 a 48 c 37 g 41 t

Query Match 100.0%; Score 23; DB 10; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGACCT 23  
 |||  
 Db 115 GAGTGTGATGTTGAGAGACCT 93

RESULT 2  
 BB583295 225 bp mRNA linear EST 30-NOV-2000  
 LOCUS BB583295 RIKEN full-length enriched, adult male cecum Mus musculus  
 DEFINITION cDNA clone 9130401H09 5', mRNA sequence.  
 ACCESSION BB583295  
 VERSION BB583295.1 GI:11479839  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.

REFERENCE  
 1 (bases 1 to 225)  
 Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.

#### AUTHORS

Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodojima, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno, H., Kusabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toyota, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M., and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Aizawa, K. et al. 2000)  
 Unpublished (2000)

#### TITLE JOURNAL COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsutsumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gs.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Thermostabilization and thermocatalysis of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)  
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

#### FEATURES

##### source

Location/Qualifiers  
 1. 225  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_1ib="9130401H09"  
 /clone\_1ib="RIKEN full-length enriched, adult male cecum"  
 /sex="male"  
 /tissue\_type="cecum"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGACTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTTCAGATTAAATTAATTCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 63 a 61 c 44 g 57 t

Query Match 100.0%; Score 23; DB 10; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGACCT 23  
 |||  
 |||

```

FEATURES
source
location/Qualifiers
1..235
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9130405818"
/clone_lib="RIKEN full-length enriched, adult male cecum"
/sex="male"
/tissue_type="cecum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site:1: SalI; Site:2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGAGAGATCCACAGACGCTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-triPPER. cDNA went through one round of normalization

```

URL: [http://genome.riken.go.jp](http://genome.gsc.riken.go.jp/Carninci.P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Thermostabilization and thermocactivation of thermostable enzymes by tritose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Itawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)</p><p>Please visit our web site (<a href=)) for further details.

Location/Qualifiers
1..244
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9030202p15"

BASE COUNT	83 a	63 c	53 g	45 t
ORIGIN	BamHI"			
Query Match	100.0%	Score 23:	DB 10:	Length 244:
Best Local Similarity	100.0%	Pred. NO. 23:		
Matches 23:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
Oy	1	GAGTCGTGATGTTGAGAGACCT	23	
Db	107	GAGTCGTGATGTTGAGAGACCT	85	
RESULT 5				
LOCUS	BB846624	284 bp	mRNA	linear
DEFINITION	BB846624 RIKEN full-length enriched, adult male kidney Mus musculus			
ACCESSION	CDNA clone F530003J06 5', mRNA sequence.			
VERSION	BB846624			
KEYWORDS	BB846624.1 GI:17084999			
SOURCE	EST.			
ORGANISM	house mouse.			
REFERENCE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 284)			
REFERENCE	1 (bases 1 to 284)			
AUTHORS	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hizamoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,P., Sato,K., Shibata,K., Shingawa,A., Shitaka,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanabe,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.			
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Yoshinhide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: res@gs.c.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)			

FEATURES	SOURCE
location/Qualifiers	1..284
/organism="Mus musculus"	
/db_xref="taxon:10090"	
/clone="P53003L06"	
/clone_id="Riken full-length enriched, adult male kidney"	
/sex="male"	
/tissue_type="kidney"	
/dev_stage="adult"	
/lab_host="SOLR"	
/note="Site_1: XhoI; Site_2: SctI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGCGCGCCCACTGAGTTTCTTTTCTTTTCTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGGATGCCAGAGGCTCATTTATTTATTTAAACCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."	
BASE COUNT	67 a 88 c 62 g 67 t
ORIGIN	
Query Match	100.0%; Score 23; DB 10; Length 284;
Best Local Similarity	100.0%; Prid. No. 24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 GAGTGTGATGTTGAGAGCCT 23 
Db	107 GAGTGTGATGTTGAGAGCCT 85
RESULT 6	
BB871941/c	354 bp mRNA linear EST 27-NOV-2001
LOCUS	BB871941 RIKEN full-length enriched adult male intestinal mucosa
DEFINITION	Mus musculus cDNA clone G63003J120 5', mRNA sequence.
ACCESSION	BB871941
VERSION	BB871941.1 GI:17118151
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 354) Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazune,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akhiba,S., Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A., Yamamatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Encyclopedia of Mouse full-length cDNAs (Akimura,T., et al.

2001)  
Unpublished (2001)  
Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216

URL: <http://genome.gsc.riken.go.jp/>  
Email: [genome-ressgsc.riken.go.jp](mailto:genome-ressgsc.riken.go.jp)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Wajiki, K., Fujiyama, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Matsui, K., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, I., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

Komono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001). Please visit our web site (<http://genome.ysc.riken.go.jp>) for further details.

## Location/Qualifiers

```
1..354
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="G650033120"
/clone_id="RKEN full-length enriched, adult male
intestinal mucosa"
/sex="male"
```

```

/issue-type="intestinal mucosa"
/dev-stage="adult"
/notice-pooled tissues: (issue-type=cerebellum,
dev-stage=16 days neonate, sex=mixed),
(issue-type=cerebellum, dev-stage=0 day neonate,
sex=mixed), (issue-type=hippocampus, dev-stage=adult,
sex=male), (issue-type=whole body, dev-stage=9 days
embryo, sex=mixed), (issue-type=lung, dev-stage=13 days

```

BASE COUNT	85 a	107 c	85 g	77 t
ORIGIN				

Query Match	100.08;	Score 23;	DB 10;	Length 354;
Best Local Similarity	100.08;	Pred. No. 26;		
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	GAGTGGTGATGTTGAGAGAGCCT	23
Db	114	GAGTGGTGATGTTGAGAGAGCCT	92

LOCUS	414 bp	mrna	linear	EST 26-NOV-2001
DEFINITION	BB848113	BB848113	RIKEN full-length enriched, adult male kidney	Mus musculus
	cdna clone F530208119	5'	mrna sequence.	

ACCESSION	BB848113
VERSION	BB848113.1
	GI:17089567

house mouse.  
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

1 (bases 1 to 414).  
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., N. Hiramoto, K. Hirooka, T. Hirozane, T. Imochi, K. Ishii, Y. Ito, M. Kawai, J. Kojima, Y. Konno, H. Kouda, M. Matsuyama, T. Nakamura, M. Nishi, K. Nomura, K. Numasaki, R. Okazaki, Y. Okido, T. Saito, R. Sakai, C. Sakai, K. Sakazume, N. Sasaki, D. Sato, K. Shibata, K. Shinagawa, A. Shiraki, T. Sogabe, Y. Suzuki, H. Tagawa, A. Tanahashi, F. Takaku, Akahira, S., Tanaka, T., Tomaru, A., Toyata, A., Watanishi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akiyura, T., et al. (2001))

Unpublished (2001)  
Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resescg.riken.go.jp/  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtriction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Wajal, K., Fujiyama, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Watanabe, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

© mouse tissues.

**source**

```
Location/Qualifiers
1..414
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="F530208119"
/clone_1fp="RKEN full-length, adult male kidney"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
```

/note="Site": XhoI; site 2 sat): cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genetic Sciences Center and genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGACAGAGAGCGCCCACTCAGACTTTTATTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGACAGAGAGAGCGCCCAAGCTCATTTAATTAAATTAACCCCCCCCC 3'], cDNA was cleaved with XhoI and SstI."

BASE COUNT	102 a	118 c	106 g	88 t
ORIGIN				

Query Match	100.0%;	Score 23;	DB 10;	Length 414;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 23; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY 1 GAGTGGTGA TGTGAGAGAGCCCT 23

[illegible]

LOCUS	637 bp	mRNA	linear	EST 26-Oct-2000	
DEFINITION	B8661515	musculus	musculus	musculus	
ACCESSION	B8661515	musculus	musculus	musculus	
VERSION	B8661515.1	GI:16495294			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota: Metazoa; Chordata: Craniata, Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 637)				
AUTHORS	Atkawa,T., Carlinici,P., Fukuda,S., Furuno,M., Hanaaki,T., Hara,A., Himoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.				
TITLE	RIKEN Mouse ESTs (Atkawa,T., et al. 2001)				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Yoshinobu Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shohiro-cho, Tsukuba-shi, Ibaraki, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/ Carlinici,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujisawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carlinici,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.				
FEATURES	Location/Qualifiers				
SOURCE	1. 637				
	/organism="Mus musculus"				
	/db_xref="taxon:10090"				
	/clone="D630043D01"				
	/clone.lib="RIKEN full-length enriched, 0 day neonate kidney"				
	/tissue="kidney"				
	/dev_stage="0 day neonate"				
	/lab_host="DH10B"				
	/note="Site.1: Sali; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGCGCCGACACTGAGTGTGTGTGTGTGTGT 3'], cDNA was prepared by using triethanolamine thermo-activated reverse				



ACCESSION	BF785145	
VERSION	BF785145.1	GI:12090181
KEYWORDS	EST.	
SOURCE	house mouse.	

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 764)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM9845 row: 1 column: 08  
High quality sequence stop: 677.  
Location/Qualifiers  
1. 764  
/organism="Mus musculus"  
/db\_xref="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4236751"  
/clone\_lib="NCI\_CGAP\_Kid14"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.75 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library. 1"  
BASE COUNT 193 a 192 c 236 g 143 t  
ORIGIN  
Query Match 100.0%; Score 23; DB 12; Length 764;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGTGTGATGTTGAGAGACCT 23  
|||||  
Db 88 GAGTGTGATGTTGAGAGACCT 66  
RESULT 13  
B1100945/c 856 bp mRNA linear EST 26-JUN-2001  
LOCUS 602886269F1 NCI\_CGAP\_Kid14 Mus musculus CDNA clone IMAGE:5041625  
DEFINITION 5' mRNA sequence.  
ACCESSION B1100945  
VERSION B1100945.1 GI:14551838  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 856)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM1114 row: 1 column: 18  
High quality sequence stop: 617.  
Location/Qualifiers  
1. 856  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="FVB/N"

/db\_xref="taxon:10090"  
/clone="IMAGE:5041625"  
/clone\_lib="NCI\_CGAP\_Kid14"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.75 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library. 1"  
BASE COUNT 249 a 220 c 249 g 138 t  
ORIGIN  
Query Match 100.0%; Score 23; DB 13; Length 856;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGTGTGATGTTGAGAGACCT 23  
|||||  
Db 75 GAGTGTGATGTTGAGAGACCT 53  
RESULT 14  
B1100201/c 879 bp mRNA linear EST 26-JUN-2001  
LOCUS 602885242F1 NCI\_CGAP\_Kid14 Mus musculus CDNA clone IMAGE:5040346  
DEFINITION 5' mRNA sequence.  
ACCESSION B1100201  
VERSION B1100201.1 GI:14551094  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 879)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM1111 row: 6 column: 11  
High quality sequence start: 4  
High quality sequence stop: 653.  
Location/Qualifiers  
1. 879  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5040346"  
/clone\_lib="NCI\_CGAP\_Kid14"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.75 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library. 1"  
BASE COUNT 227 a 256 c 243 g 153 t  
ORIGIN  
Query Match 100.0%; Score 23; DB 13; Length 879;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGTGTGATGTTGAGAGACCT 23  
|||||  
Db 113 GAGTGTGATGTTGAGAGACCT 91  
RESULT 15

BF781298/c 897 bp mRNA linear EST 12-JAN-2001  
 LOCUS 602106974F1 NCI\_CGAP\_Kid14 Mus musculus cDNA clone IMAGE:4235106  
 DEFINITION 5' mRNA sequence.  
 ACCESSION BF781298  
 VERSION BF781298.1 GI:12086334  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 897)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM9841 row: d column: 19  
 High quality sequence stop: 706.

FEATURES  
 Source location/Qualifiers  
 1..897

/organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4235106"  
 /clone\_lib="NCI\_CGAP\_Kid14"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Organ: Kidney; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library. 1"  
 BASE COUNT 219 a 244 c 277 g 157 t  
 ORIGIN

Query Match 100.0%; Score 23; DB 12; Length 897;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGGTGTGAGAGCCT 23  
 ||||||||||||||||||  
 Db 91 GAGTGGTGTGAGAGCCT 69

Search completed: January 24, 2003, 18:53:16  
 Job time : 33.4042 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:33 : Search time 46.6937 Seconds  
(without alignments)  
12384.474 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23  
Sequence: 1 gagtggatgttgagagagcct 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main:\*

```

1: /cgn2_6/ptodata/2/pna/US0600.COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06.COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07.COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US080.COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081.COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082.COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083.COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084.COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085.COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086.COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087.COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088.COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089.COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090.COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091.COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092.COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093.COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094.COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A.COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B.COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C.COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D.COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A.COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B.COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C.COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096D.COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096E.COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097A.COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097B.COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097C.COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US098A.COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US098B.COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US098C.COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US099A.COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US099B.COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US099C.COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US099D.COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US100A.COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US100B.COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US101A.COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US101B.COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US102A.COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US102B.COMB.seq.*

```

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	US-09-877-935-2	Sequence 2, Appl1
2	23	100.0	380	US-10-060-066-210	Sequence 210, Ap
3	23	100.0	3049	US-60-360-207-15646	Sequence 15646, A
4	21.4	93.0	98	US-08-242-057-4	Sequence 4, Appl1
5	21.4	93.0	98	US-08-242-057B-4	Sequence 4, Appl1
6	21.4	93.0	98	US-08-467-454A-4	Sequence 4, Appl1
7	21.4	93.0	98	US-08-467-454B-4	Sequence 4, Appl1
8	21.4	93.0	417	US-09-716-473-867	Sequence 867, App
9	21.4	93.0	1991	US-08-242-057B-1	Sequence 1, Appl1
10	21.4	93.0	1991	US-08-467-454B-1	Sequence 1, Appl1
11	21.4	93.0	3467	US-09-205-070-4233	Sequence 4233, Ap
12	21.4	93.0	3467	US-09-340-623-4233	Sequence 4233, Ap
13	21.4	93.0	3467	US-09-898-888-4233	Sequence 4233, Ap
14	21.4	93.0	3467	US-09-898-888A-4233	Sequence 4233, Ap
15	21.4	93.0	3487	US-09-496-914A-8145	Sequence 8145, Ap
16	21.4	93.0	3487	US-09-560-815A-8145	Sequence 8145, Ap
17	21.4	93.0	4519	US-09-359-922-2625	Sequence 2625, Ap
18	21.4	93.0	4519	US-09-359-922-2625	Sequence 2625, Ap
19	21.4	93.0	4519	US-09-919-002-2625	Sequence 2625, Ap
20	21.4	93.0	5219	US-60-172-373-10168	Sequence 10168, A
21	21.4	93.0	5287	US-60-276-258-12273	Sequence 12273, A

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

```
C 22 21.4 93.0 5680 23 US-09-606-776-4585 Sequence 4585, Ap
C 23 21.4 93.0 5680 25 US-09-652-126-9168 Sequence 9168, Ap
C 24 21.4 93.0 5680 25 US-09-652-915-10128 Sequence 10128, A
C 25 21.4 93.0 5680 25 US-09-652-916-9200 Sequence 9200, Ap
C 26 21.4 93.0 5680 27 US-09-698-010-12819 Sequence 12819, A
C 27 21.4 93.0 5680 28 US-09-716-473-2399 Sequence 2399, Ap
C 28 21.4 93.0 5680 29 US-09-717-350-5084 Sequence 5084, Ap
C 29 21.4 93.0 5680 29 US-09-726-174-4654 Sequence 4654, Ap
C 30 21.4 93.0 5680 29 US-09-726-174-4654 Sequence 4654, Ap
C 31 19.8 86.1 506 69 US-60-252-833-30632 Sequence 30632, A
C 32 19.8 86.1 579 69 US-60-252-833-30632 Sequence 30632, A
C 33 19.8 86.1 672 69 US-60-252-833-30108 Sequence 30108, A
C 34 19.8 86.1 672 69 US-60-252-833-30108 Sequence 30108, A
C 35 18.4 80.0 368 69 US-60-253-457-29594 Sequence 29594, A
C 36 18.4 80.0 407 25 US-09-654-617-336155 Sequence 336155,
C 37 18.4 80.0 407 27 US-09-684-016-336155 Sequence 336155,
C 38 18.4 80.0 407 27 US-09-684-016-336155 Sequence 336155,
C 39 18.4 80.0 407 60 US-60-162-747-3797 Sequence 3797, Ap
C 40 18.4 80.0 591 69 US-60-253-456-18428 Sequence 18428, A
C 41 18.4 80.0 1674 42 US-10-219-999-29253 Sequence 29253, A
C 42 18.4 80.0 1707 33 US-09-865-439A-20985 Sequence 20985, A
C 43 18.4 80.0 2522 33 US-09-867-716-17994 Sequence 17994, A
C 44 18.4 80.0 2522 33 US-09-886-492-17994 Sequence 17994, A
C 45 18.4 80.0 14702 19 US-09-528-237A-2208 Sequence 2208, Ap
```

## ALIGNMENTS

```
RESULT 1
US-09-877-935-2
; Sequence 2, Application US/09877935
; GENERAL INFORMATION:
; APPLICANT: Pinto, Daniel
; APPLICANT: Robine, Sylvie
; APPLICANT: Ualser, Frederic
; APPLICANT: Louvard, Daniel
; TITLE OF INVENTION: REGULATORY SEQUENCES OF THE MOUSE VILIN GENE - USE IN TRANSGENES
; FILE REFERENCE: 13294-002001
; CURRENT APPLICATION NUMBER: US/09/877, 935
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: PCT/EP 98/08009
; PRIOR FILING DATE: 1998-12-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: oligonucleotide
US-09-877-935-2
```

```
Query Match 100.0%; Score 23; DB 33; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 GAGTGTGATGTGAGAGCCT 23
Db 1 GAGTGTGATGTGAGAGCCT 23
```

```
RESULT 2
US-10-060-066-210/C
; Sequence 210, Application US/10060066
; GENERAL INFORMATION:
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NOVEL MUTATED MAMMALIAN CELLS AND
; TITLE OF INVENTION: ANIMALS
; FILE REFERENCE: LEX-0304-USA
; CURRENT APPLICATION NUMBER: US/10/060, 066
```

```
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 60/265,574
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(380)
; OTHER INFORMATION: n = A,T,C or G
US-10-060-066-210
```

```
Query Match 100.0%; Score 23; DB 39; Length 380;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 GAGTGTGATGTGAGAGCCT 23
Db 65 GAGTGTGATGTGAGAGCCT 43
```

```
RESULT 3
US-60-360-207-15646/C
; Sequence 15646, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CLO01321
; CURRENT APPLICATION NUMBER: US/60/360, 207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 15646
; LENGTH: 3049
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-15646
```

```
Query Match 100.0%; Score 23; DB 80; Length 3049;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 GAGTGTGATGTGAGAGCCT 23
Db 91 GAGTGTGATGTGAGAGCCT 69
```

```
RESULT 4
US-08-242-057-4/C
; Sequence 4, Application US/08242057
; GENERAL INFORMATION:
; APPLICANT: Pringault, Eric
; APPLICANT: Robine, Sylvie
; APPLICANT: Huelt, Christian
; APPLICANT: Babinet, Charles
; APPLICANT: Louvard, Daniel
; TITLE OF INVENTION: Villin Gene Promoter Sequence and Its
; TITLE OF INVENTION: Use in Vectors, Transformed Mammalian Cell Lines,
; TITLE OF INVENTION: Transgenic Animals, and Cell Lines Derived from the
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```







FILING DATE: 13-MAY-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,905  
FILING DATE: 29-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, Leslie A.  
REGISTRATION NUMBER: 34,872  
REFERENCE/DOCKET NUMBER: 03495.0096-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1991 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-242-057B-1

Query Match 93.0%; Score 21.4; DB 6; Length 1991;  
Best Local Similarity 95.7%; Pred. No. 28;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23  
Db 1974 GGGTGTGATGTGAGAGCCT 1952

RESULT 10  
US-08-467-454B-1/C  
Sequence 1, Application US/08467454B  
GENERAL INFORMATION:  
APPLICANT: PRINGAULT, Eric  
APPLICANT: ROBIN, Sylvie  
APPLICANT: HUET, Christian  
APPLICANT: BABINET, Charles  
APPLICANT: LOUYARD, Daniel  
TITLE OF INVENTION: VILIN GENE PROMOTER SEQUENCE AND ITS  
TITLE OF INVENTION: USE IN VECTORS, TRANSFORMED MAMMALIAN CELL LINES,  
TITLE OF INVENTION: TRANSGENIC ANIMALS, AND CELL LINES DERIVED FROM THE  
TITLE OF INVENTION: ANIMALS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,454B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/242,057  
FILING DATE: 13-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/121,454  
FILING DATE: 16-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,905  
FILING DATE: 29-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: McDonnell, Leslie A.  
REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 03495.0096-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1991 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-467-454B-1

Query Match 93.0%; Score 21.4; DB 8; Length 1991;  
Best Local Similarity 95.7%; Pred. No. 28;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23  
Db 1974 GGGTGTGATGTGAGAGCCT 1952

RESULT 11  
US-09-205-070-4233/C  
Sequence 4233, Application US/09205070  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA  
FILE REFERENCE: 20411-748  
CURRENT APPLICATION NUMBER: US/09/205,070  
CURRENT FILING DATE: 1998-12-03  
NUMBER OF SEQ ID NOS: 45207  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4233  
LENGTH: 3467  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-205-070-4233

Query Match 93.0%; Score 21.4; DB 16; Length 3467;  
Best Local Similarity 95.7%; Pred. No. 30;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23  
Db 73 GGGTGTGATGTGAGAGCCT 51

RESULT 12  
US-09-340-623-4233/C  
Sequence 4233, Application US/09340623  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA  
FILE REFERENCE: 20411-748CON1  
CURRENT APPLICATION NUMBER: US/09/340,623  
CURRENT FILING DATE: 1999-06-28  
EARLIER APPLICATION NUMBER: US 09/205,070  
EARLIER FILING DATE: 1998-12-03  
NUMBER OF SEQ ID NOS: 45207  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4233  
LENGTH: 3467  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-340-623-4233

Query Match 93.0%; Score 21.4; DB 17; Length 3467;  
Best Local Similarity 95.7%; Pred. No. 30;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	GAGTGGTGATGTTGAGAGAGCCT	23
Db	73	GGGTGGTGATGTTGAGAGAGACCT	51

```

RESULT 13
US-09-898-888-4233/c
Sequence 4233. Application US/0989888
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
TITLE OF INVENTION: LIBRARIES
FILE REFERENCE: 20411-748CON1
CURRENT APPLICATION NUMBER: US/09/898,888
CURRENT FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: 09/340,623
PRIOR FILING DATE: 1999-06-28
NUMBER OF SEQ ID NOS: 45207
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4233
LENGTH: 3467
TYPE: DNA
ORGANISM: Homo sapiens
US-09-898-888-4233

```

Query Match	93.0%	Score 21.4	DB 33	Length 3467
Best Local Similarity	95.7%	Pred. No. 30		
Matches 22: Conservative	0:	Mismatches 1:	Indels 0:	Gaps 0:
Oy	1	GAGTGTGATGCTTGGAGAGGCT	23	
Db	73	GGGTGTGATGCTTGGAGAGGCT	51	

```

RESULT 14
US-09-898-888A-4233/C
: Sequence 4233, Application US/09898888A
: GENERAL INFORMATION:
:   APPLICANT: Hyseq, Inc.
:   TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
:   TITLE OF INVENTION: LIBRARIES
:   FILE REFERENCE: 20411-748CON1
:   CURRENT APPLICATION NUMBER: US/09/898,888A
:   CURRENT FILING DATE: 2001-07-30
:   PRIOR APPLICATION NUMBER: US/09/340,623
:   PRIOR FILING DATE: 1999-06-28
:   PRIOR APPLICATION NUMBER: US 09/205,070
:   PRIOR FILING DATE: 1998-12-03
:   NUMBER OF SEQ ID NOS: 45207
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 4233
:   LENGTH: 3467
:   TYPE: DNA
:   ORGANISM: Homo sapiens
US-09-898-888A-4233

```

Query Match	93.0%	Score 21.4	DB 33	Length 3467
Best Local Similarity	95.7%	Pred. No. 30		
Matches 22	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1	GAGTGTGATGTTTGAGAGACCT	23	
Db	73	GAGTGTGATGTTTGAGAGACCT	51	

```

RESULT15
US-09-496-914A-8145/c
; Sequence 8145, Application US/09496914A
; GENERAL INFORMATION:
; APPLICANT: Tang, Yuanhua T.
; APPLICANT: Tillinghast, John
; APPLICANT: Sinku, Ankura
; APPLICANT: Liu, Chenghua

```

```

APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Novel Configs Obtained
TITLE OF INVENTION: From Various Libraries
FILE REFERENCE: 787
CURRENT APPLICATION NUMBER: US/09/496, 914A
CURRENT FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/353, 690
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 09/034,341
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 09/045, 400
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: US 09/321, 214
PRIOR FILING DATE: 1999-05-26
PRIOR APPLICATION NUMBER: US 09/131, 598
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: US 09/431, 517
PRIOR FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: US 09/328, 351
PRIOR FILING DATE: 1999-06-04
PRIOR APPLICATION NUMBER: US 09/332, 782
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/235, 076
PRIOR FILING DATE: 1999-01-20
PRIOR APPLICATION NUMBER: US 09/234, 611
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: US 09/346, 956
PRIOR FILING DATE: 1999-07-02
PRIOR APPLICATION NUMBER: US 09/362, 510
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: US 09/240, 371
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: US 09/248, 797
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 09/271, 490
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: US 09/293, 972
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: US 09/274, 861
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/125, 453
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: US 60/126, 605
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: US 09/306, 350
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/399, 720
PRIOR FILING DATE: 1999-09-21
PRIOR APPLICATION NUMBER: US 09/404, 284
PRIOR FILING DATE: 1999-09-21
PRIOR APPLICATION NUMBER: US 09/465, 877
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
PRIOR FILING DATE: 1999-01-19
NUMBER OF SEQ ID NOS: 10410
SOFTWARE: pt_CT_genes Version 1.02
SEQ ID NO 8145
LENGTH: 3487
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44) ..(2526)
OTHER INFORMATION: similar to g1412172 in the genepept database release 114
US-09-496-914A-8145
OTHER INFORMATION: Run with FASTxY 3.3f000, default parameters

```

Query Match	93.0%	Score 21.4	DB 18	Length 3487
Best Local Similarity	95.7%	Pred. No. 30		
Matches 22: Conservative	0	Mismatches	1	Indels 0
				Gaps 0
QY	1	GAGGTGATGTTGAGAGCGCT	23	

Tue Jan 28 08:20:21 2003

us-09-877-935-2.rnpm

Page 7

Db 92 GGGTGGTGATGTTGAGAGAGCCCT 70

Search completed: January 25, 2003, 07:47:10  
Job time : 48.6937 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:31 : Search time 4.27201 Seconds  
(without alignments)  
12124.490 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23

Sequence: 1 gagtggatgtgttgagagagcct 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

N.Geneseq\_101002:\*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	AAA48516	Murine villin gene
2	23	100.0	23	AAA49254	Mouse villin gene
3	21.4	93.0	1990	AAO25816	Villin gene promot
4	21.4	93.0	2702	AAO04681	Human villin gene
5	18.4	80.0	4286	AA18079	Heml gene. H2-1 b
6	18.4	80.0	4286	AA18079	Heml gene. H2-1 b
7	18.2	79.1	1663	ABN88181	G protein-coupled
8	17.4	75.7	1475	AAV00527	HIV-1 breakthrough
9	17.4	75.7	1475	AAV00528	HIV-1 breakthrough

C 10	17.2	74.8	538	23	ABV54721	Human prostate exp
C 11	17.2	74.8	963	21	AAC46212	Arabidopsis thalia
C 12	17.2	74.8	1054	21	AAC45514	Arabidopsis thalia
C 13	17.2	74.8	1056	21	AAC37611	Arabidopsis thalia
C 14	17.2	74.8	2884	21	AAV79490	Eucalyptus grandis
C 15	17.2	74.8	3600	21	AAV79657	Eucalyptus grandis
C 16	17.2	74.8	21423	22	AAU36470	Human musculoskele
C 17	17.2	74.8	22073	22	AAU04663	Human reproductive
C 18	17.2	74.8	22073	23	ABL97570	Human testicular a
C 19	16.8	73.0	4774	22	AAK64331	Human immune/haema
C 20	16.8	73.0	2955	24	ABL91678	Human polynucleoti
C 21	16.8	73.0	4281	14	AAO53471	elk CDNA. Rattus
C 22	16.8	73.0	7924	24	ABK40069	Human chemically p
C 23	16.8	73.0	7924	24	ABL34138	Human immune syste
C 24	16.8	73.0	15650	22	ABR15903	Human nervous syst
C 25	16.8	73.0	15650	22	ABR16564	Human nervous syst
C 26	16.8	73.0	143068	21	AAE21105	Human low adenosin
C 27	16.8	73.0	143068	21	AAE21272	Human low adenosin
C 28	16.8	73.0	143068	21	AAA34983	Human adenosine re
C 29	16.8	73.0	143068	21	AAA35150	Human adenosine re
C 30	16.8	73.0	143068	24	ABL68124	Ovary cancer relat
C 31	16.8	73.0	149412	21	AAA35151	Human adenosine re
C 32	16.8	73.0	152740	21	AAE21273	Human low adenosin
C 33	16.8	73.0	2365589	24	ABA90521	Genomic sequence o
C 34	16.6	72.2	321	22	AAK87824	Human digestive sy
C 35	16.6	72.2	387	24	ABN24942	Human ORF polynuc
C 36	16.6	72.2	600	21	AAE08568	Fusarium venenatum
C 37	16.6	72.2	602	22	AAH87718	Pepperfruit plant o
C 38	16.6	72.2	4618	23	ABK42604	Genomic sequence #
C 39	16.6	72.2	7353	24	ABN89472	Human G-protein co
C 40	16.6	72.2	12989	23	ABK42602	Genomic sequence #
C 41	16.6	72.2	24978	20	AAK60209	SEQ ID 3 of US9314
C 42	16.6	72.2	24979	21	AAA52321	Genomic DNA sequen
C 43	16.6	72.2	129722	24	ABQ88117	Human osteoblast d
C 44	16.4	71.3	904	22	AAH32619	Human secreted pro
C 45	16.4	71.3	975	23	AA594424	DNA encoding novel

#### ALIGNMENTS

RESULT 1	AAA48516	standard; DNA; 23 BP.
ID	AAA48516	
XX	AAA48516:	
AC	19-DEC-2000 (first entry)	
XX		
DT		
XX		
DE	Murine villin gene regulatory region probe #1.	
XX		
KW	Mouse; villin; regulatory region; digestive tract;	
KW	colorectal cancer mouse model; probe; ss.	
XX		
OS	Mus sp.	
XX		
PN	WO200034492-A1.	
XX		
PD	15-JUN-2000.	
XX		
PF	09-DEC-1998; 98WO-EP08009.	
XX		
PA	09-DEC-1998; 98WO-EP08009.	
XX		
PA	(CNRS ) CENT NAT RECH SCI.	
XX	(CURT-) INST CURIE.	
XX		
PI	Pinto D, Robine S, Jaisser F, Louvard D;	
XX		
DR	WPI: 2000-423433/36.	
XX		
PT	Novel nucleotide sequence derived from mouse villin gene for targeted expression of transgenes in immature and differentiated epithelial	

PT cells of intestine or urogenital tracts -  
XX  
PS Disclosure; Page 13; 54pp; English.  
XX  
CC The present sequence is a probe which was used in primer extension  
CC analysis of the murine villin gene regulatory region. It has been  
CC shown that this region directs the expression of the villin gene in the  
CC intestine and uro-genital tracts, and thus could be used in a fusion  
CC gene to direct expression of exogenous genes in these areas. This could  
CC be used, for example, to create a mouse model for colorectal cancer.  
XX  
SQ Sequence 23 BP; 5 A; 2 C; 10 G; 6 T; 0 other;  
Query Match 100.0%; Score 23; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GAGTGTGATGTTGAGAGAGCCT 23  
DB 1 GAGTGTGATGTTGAGAGAGCCT 23  
RESULT 2  
AAA9254  
ID AAA9254 standard; DNA: 23 BP.  
XX  
AC AAA9254;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Mouse villin gene oligonucleotide probe #1.  
XX  
KM Mouse; villin; intestinal epithelial cell;  
KM uro-genital tract epithelial cell; tumour; probe; ss.  
XX  
OS Mus sp.  
XX  
PN MO200034493-A2.  
XX  
PD 15-JUN-2000.  
XX  
PE 09-DEC-1999; 99WO-EP09782.  
XX  
PR 09-DEC-1998; 98WO-EP08009.  
XX  
PA (CNRS ) CENT NAT RECH SCI.  
PA (CURT-) INST CURIE.  
XX  
PI Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J;  
DR WPI; 2000-423434/36.  
XX  
PT Novel nucleotide sequence derived from mouse villin gene for targeted  
PT expression of transgenes in immature and differentiated epithelial  
PT cells of intestine or urogenital tracts -  
XX  
PS Example 1; Page 13; 52pp; English.  
XX  
CC The present sequence is an oligonucleotide probe for the murine  
CC villin gene. This gene is expressed in the epithelial cells of the  
CC intestine and uro-genital tracts. Its promoter sequence can be used in  
CC the targeted expression of exogenous genes in these places, which may,  
CC for example, be useful in the treatment of tumours. This probe was used  
CC in a primer extension assay.  
XX  
SQ Sequence 23 BP; 5 A; 2 C; 10 G; 6 T; 0 other;  
Query Match 100.0%; Score 23; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GAGTGTGATGTTGAGAGAGCCT 23  
DB 1 GAGTGTGATGTTGAGAGAGCCT 23

DB 1 GAGTGTGATGTTGAGAGAGCCT 23  
RESULT 3  
AAQ25816/c  
ID AAQ25816 standard; DNA: 1990 BP.  
XX  
AC AAQ25816;  
XX  
DT 06-JAN-1993 (first entry)  
XX  
DE Villin gene promoter.  
XX  
KM Oncogene; embryogenesis; terminal differentiation; cell lineage;  
KM carcinogenesis; metastasis; chloramphenicol acetyltransferase; CAT;  
KM human growth hormone; hGH; luciferase; beta-galactosidase; ds.  
XX  
OS Homo sapiens.  
XX  
PN CA2054149-A.  
XX  
PD 30-APR-1992.  
XX  
PE 24-OCT-1991; 91CA-2054149.  
XX  
PR 29-OCT-1990; 90US-0604905.  
XX  
PA (PRIN/) PRINGAUT E.  
XX  
PI Babinet C, Huët C, Louvard D, Pringault E, Robine S, Louvard C;  
XX  
DR WPI; 1992-235048/29.  
XX  
PT Promoter sequence from human villin gene - used in vectors linked  
PT to e.g. oncogene for prodn. of transgenic animal models of  
PT colorectal cancer  
XX  
PS Disclosure; Fig 4"; 63pp; English.  
XX  
CC This promoter sequence can be operably linked to a gene eg. a an  
CC oncogene. Vectors containing such constructs are useful in the  
CC production of animal models of eg. colorectal cancer. Villin gene  
CC expression is strictly regulated during embryogenesis and terminal  
CC differentiation and villin is a marker of digestive epithelial cell  
CC lineage. The tissue-specific expression is maintained during  
CC carcinogenesis and villin can therefore be used to characterise the  
CC colorectal origin of unknown metastasis. The promoter can also be  
CC linked to a structural gene and used in transient assay systems based  
CC on the use of fusion genes. Suitable reporter genes are  
CC chloramphenicol acetyltransferase (CAT), human growth hormone (hGH),  
CC luciferase and beta-galactosidase. This promoter is also useful in  
CC the investigation of the molecular genetics involved in digestive  
CC organogenesis and may be used as a probe for the detection of  
CC a nucleotide sequence in a biological material.  
XX  
SQ Sequence 1990 BP; 549 A; 505 C; 546 G; 390 T; 0 other;  
Query Match 93.0%; Score 21.4; DB 13; Length 1990;  
Best Local Similarity 95.7%; Pred. No. 2.4;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GAGTGTGATGTTGAGAGAGCCT 23  
DB 1973 GAGTGTGATGTTGAGAGAGCCT 1951  
RESULT 4  
AAQ04681/c  
ID AAQ04681 standard; DNA: 2702 BP.  
XX  
AC AAQ04681;  
XX  
DT 20-SEP-1990 (first entry)

```

XX DE Human villin gene fragments for probes and primers.
XX XX
XX KW Human villin gene; amplification primers; probes; tumour diagnosis;
XX KW RFLP; ss
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 25..2505
XX FT /*tag= a
XX FT /product= human villin
XX PN EP367666-A.
XX PD 09-MAY-1990.
XX PF 27-OCT-1989; 89EP-0402984.
XX PR 28-OCT-1988; 88FR-0014208; EP-402984.
XX PA (INSP ) INST PASTEUR.
XX P1 Arpin M, Pringault E, Garcia A, Louvard D;
XX DR WPI: 1990-141549/19.
XX DR P-PSDB: AAR05049.
XX PT New nucleic acid fragments corresponding to villin gene - useful as
XX PT diagnostic amplification primers or hybridisation probes.
XX PS Disclosure: ; p: French.
XX XX
XX CC Any nucleic acid fragment comprising a sequence of 8-40 ,pref.20-40,
XX CC nucleotides which is present in the human villin gene, or is exactly
XX CC complementary to such a sequence, or is present in RNA fragment corresp.
XX CC to such a sequence is useful as primer for amplifying villin-encoding DNA
XX CC or the corresp. mRNA in biological samples, esp. in the diagnosis of
XX CC gastrointestinal tumours or metastases, and as probes for detecting
XX CC RFLPs of the human genome.
XX SQ Sequence 2702 BP; 681 A; 719 C; 778 G; 524 T; 0 other;
XX
XX Query Match 93.0%; Score 21.4; DB 11; Length 2702;
XX Best Local Similarity 95.7%; Pred. No. 2.5;
XX Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGTGTGATGTTGAGAGAGCCT 23
Db 73 GGGTGGTGTGTTGAGAGAGCCT 51

```

```

XX XX
XX PR 27-MAY-1994; 94JP-0115597.
XX XX
XX PA (NASC-) NAT SCI COUNCIL.
XX XX
XX DR WPI: 1996-154854/16.
XX XX
XX PT Prevention of virus infections by induction of apoptosis - uses
XX PT anti-virus gene hem1 (persistence-associated gene 1) to cause
XX PT programmed cell death upon infection
XX PS Disclosure: Page 16-17; 28pp; Japanese.
XX XX
XX CC This sequence represents the HZ-1 baculovirus hem1 gene. Hem1 is also
XX CC known as the persistence-associated gene 1. This sequence can be used in
XX CC the method of the invention for prevention and/or treatment of viral
XX CC infections, by induction of earlier apoptosis in the animal host cells or
XX CC living body. This method comprises making the animal host cells
XX CC susceptible to apoptosis induction upon virus infection and then inducing
XX CC apoptosis by attack of the infectious virus on the host animal cells.
XX CC This causes cell death, followed by blocking the regeneration of the
XX CC infectious virus, thereby killing the virus in the host animal cells. In
XX CC order to carry out this method, the host cells contain an anti-virus gene
XX CC (such as this gene). This gene is preferably introduced to the cell by
XX CC genetic recombination. This method may be used for the prevention of
XX CC virus infection.
XX SQ Sequence 4286 BP; 1256 A; 791 C; 882 G; 1357 T; 0 other;
XX
XX Query Match 80.0%; Score 18.4; DB 17; Length 4286;
XX Best Local Similarity 95.0%; Pred. No. 66;
XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGTGTGATGTTGAGAGAG 20
Db 144 GAGGGGTGATGTTGAGAGAG 125

```

```

RESULT 6
AAAT6880/C
ID AAAT6880 standard; DNA; 4286 BP.
XX AC AAAT6880;
XX DF 05-AUG-1999 (first entry)
XX DE HZ-1 pag1 promoter.
XX KW HZ-1 pag1 promoter; persistence-associated gene 1; insect cell;
XX KW constitutive expression promoter; ss.
XX OS HZ-1 virus.
XX PN US5911982-A.
XX PD 15-JUN-1999.
XX PF 18-APR-1996; 96US-0634350.
XX PR 18-APR-1996; 96US-0634350.
XX PR 06-OCT-1995; 95US-0004894.
XX PR 11-OCT-1995; 95US-0005128.
XX PA (NASC-) NAT SCI COUNCIL.
XX PI Chao Y;
XX DR WPI: 1999-357167/30.
XX PT HZ-1 virus persistence-associated gene promoter
XX PS Claim 1; Fig 1b; 56pp; English.
XX XX

```

CC This sequence represents the Hx-1 persistence-associated gene 1  
 CC (Hx-1 pag1) promoter of the invention. The pag1 gene promoter is useful  
 CC in insect cells for driving constitutive expression of e.g. genes  
 CC encoding foreign proteins. The promoter of the pag1 gene is  
 CC constitutively expressed and stronger than that of the polyhedrin gene in  
 CC insect cells, enabling it to express foreign genes more strongly  
 CC e.g. lacZ and luciferase, in addition to which it can be expressed more  
 CC prominently as a short promoter.

XX  
 SQ Sequence 4286 BP; 1256 A; 792 C; 881 G; 1357 T; 0 other;

Query Match 80.0%; Score 18.4; DB 20; Length 4286;  
 Best Local Similarity 95.0%; Pred. No. 66;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GAGTGTGATGTGAGAGAG 20  
 ||| ||||| ||||| |||||  
 Db 144 GAGGTGTGATGTGAGAGAG 125

# RESULT 7

ID ABN8181 standard; DNA: 1663 BP.

AC ABN8181;

DT 13-AUG-2002 (first entry)

DE G protein-coupled receptor CGR95 nucleotide sequence.

XX RNA analysis: identification; RNA molecule: antibacterial; virucide;

KW functional; cytosolic; antisense therapy; cancer; infection;

KM G protein-coupled receptor; gene: ds.

XX Unidentified.

OS WO200224950-A2.

PN 28-MAR-2002.

PD 25-SEP-2001; 2001MO-SE02054.

XX 25-SEP-2000; 2000US-235029P.

PR (NEUR-) NEUROWICS INC.

XX Liang Z, Zhang H, Wahlestedt C;

PI MPI; 2002-404959/43.

XX Identifying accessible region (AR) of native RNA, involves selecting

PT from oligonucleotide population, an oligonucleotide binding to AR,

PT sequencing randomized portion of oligonucleotide, and identifying

PT sequence of AR

XX Example: Fig 8; 41pp; English.

PS The present invention describes a method (M1) for the single-cycle

XX identification of an accessible region (AR) of native RNA (I). The method

CC comprises providing an in vitro reaction mixture comprising (i) and a

CC population of oligonucleotides (II), each having a randomized portion

CC that can bind to a complementary AR of (I), if present, selecting a (II)

CC that binds to an AR, sequencing the randomized portion of (II), and

CC identifying the nucleotide sequence of the AR. An AR can have virucide,

CC antibacterial, fungicide and cytostatic activities, and can be used in

CC an AR of a native RNA preferably mRNA. Identifying an AR of mRNA is

CC useful for manufacturing an antisense oligonucleotide for the

CC downregulation of expression of an mRNA molecule which involves

CC identifying an AR on an mRNA using the method and synthesizing an

CC oligonucleotide complementary to AR. (M1) is useful for making an

CC antisense oligonucleotide which involves identifying an AR of a native

CC RNA by (M1) and synthesizing the antisense oligonucleotide that is

CC complementary to the AR. The antisense oligonucleotides are useful for

CC treating disorders associated with aberrant gene expression, such as

CC cancer and disorders associated with expression of foreign genes such

CC as infection with bacterial, viral or fungal pathogen. The present

CC sequence represents an oligonucleotide which is used in the

CC exemplification of the present invention.

XX  
 SQ Sequence 1663 BP; 341 A; 461 C; 440 G; 421 T; 0 other;

Query Match 79.1%; Score 18.2; DB 24; Length 1663;  
 Best Local Similarity 87.0%; Pred. No. 71;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GAGTGTGATGTGAGAGAGCCT 23  
 ||||| ||||| ||||| |||||  
 Db 106 GAGTGTGATGTAAAGACAGCCT 84

DE HIV-1 breakthrough isolate clone C10.5 gp120 polypeptide DNA.

XX HIV-1; envelope protein; gp120; MN-rp120; vaccine; AIDS; ss.

KW Human immunodeficiency virus type 1.

XX Key Location/Qualifiers

FT CDS 2..1450

FT /tag= a

XX WO9801564-A1.

PN 15-JAN-1998.

PD 03-JUL-1997; 97MO-US09690.

XX 08-JUL-1996; 96US-0676737.

PR (GETH ) GENENTECH INC.

XX Berman PW;

PI MPI; 1998-101059/09.

XX P-PSDB; AAM37064.

XX Human immunodeficiency virus gp120 sequences from vaccine

PT breakthrough strains - useful in providing added protection against

PT HIV not provided by original vaccine

XX Claim 6: Page 133-135; 193pp; English.

PS Clone C10.5 encodes a gp120 envelope polypeptide (see AAM37064) of

XX a breakthrough isolate of HIV-1 obtained from an individual

CC infected with HIV-1 through high risk activity while participating

CC in Phase I or Phase 2 trials of MN-rp120, a candidate recombinant

CC gp120 HIV-1 vaccine. Nucleotide sequences (see AAV00517-30) for

CC gp120 polypeptides (see AAM37054-67) were obtained from 2 clones of

CC each of 7 breakthrough isolates. The envelope glycoprotein genes

CC were obtained from proviral DNA using PCR (see AAV00533-38). All 7

CC envelope glycoproteins possess sequences typical of subtype (clade)

CC B viruses. The overall homology with MN-rp120 is 69-80%. Use of

CC the gp120 polypeptides from one or more of the breakthrough

CC isolates, usually together with MN-rp120, can provide protection

CC against HIV strains that are sufficiently different from the

CC vaccine strain (e.g. MN-rp120) that the vaccine does not confer

CC protection against those strains.



SO Sequence 1475 BP; 568 A; 240 C; 314 G; 353 T; 0 other;

Query Match 75.7%; Score 17.4; DB 19; Length 1475;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GTGGTGAATGTTGAGAGAC 21  
|||||  
DB 375 GTGGTGAATGTTGAGAGAC 357

RESULT 9  
AAV00528/C  
ID AAV00528 standard; DNA; 1475 BP.

AC AAV00528;

DT 20-JUL-1998 (first entry)

DE HIV-1 breakthrough isolate clone C10.7 gp120 polypeptide DNA.

KM HIV-1; envelope protein; gp120; MN-rgp120; vaccine; AIDS; ss.

OS Human immunodeficiency virus type 1.

FT Key Location/Qualifiers

FT CDS 2..1465

PN MO9801564-A1.

PD 15-JAN-1998.

PF 03-JUL-1997; 97WO-US09690.

PR 08-JUL-1996; 96US-0676737.

PA (GENE) GENENTECH INC.

PI Berman PW;

DR WPI; 1998-101059/09.

DR P-PSDB; AAM37065.

PT Human immunodeficiency virus gp120 sequences from vaccine

PS Claim 6; Page 137-139; 193pp; English.

CC Clone C10.7 encodes a gp120 envelope polypeptide (see AAM37065) of  
CC a breakthrough isolate of HIV-1 obtained from an individual  
CC infected with HIV-1 through high risk activity while participating  
CC in Phase I or Phase 2 trials of MN-rgp120, a candidate recombinant  
CC gp120 HIV-1 vaccine. Nucleotide sequences (see AAV00517-30) for  
CC gp120 polypeptides (see AAM37054-67) were obtained from 2 clones of  
CC each of 7 breakthrough isolates. The envelope glycoprotein genes of  
CC were obtained from proviral DNA using PCR (see AAV00533-38). All 7  
CC envelope glycoproteins possess sequences typical of subtype (clade)  
CC B viruses. The overall homology with MN-rgp120 is 69-80%. Use of  
CC the gp120 polypeptides from one or more of the breakthrough  
CC isolates, usually together with MN-rgp120, can provide protection  
CC against HIV strains that are sufficiently different from the  
CC vaccine strain (e.g. MN-rgp120) that the vaccine does not confer  
CC protection against those strains.

SO Sequence 1475 BP; 566 A; 240 C; 316 G; 353 T; 0 other;

Query Match 75.7%; Score 17.4; DB 19; Length 1475;  
Best Local Similarity 94.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GTGGTGAATGTTGAGAGAC 21

DB 375 GTGGTGAATGTTGAGAGAC 357

RESULT 10  
ABV54721/C  
ID ABV54721 standard; cDNA; 538 BP.

AC ABV54721;

DT 17-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 54712.

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KM pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

PR 17-FEB-2000; 2000US-183119P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer

PS Claim 1; Page 10575; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the  
CC specification, or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

SO Sequence 538 BP; 113 A; 134 C; 139 G; 152 T; 0 other;

Query Match 74.8%; Score 17.2; DB 23; Length 538;  
Best Local Similarity 86.4%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ACTGCTGAATGTTGAGAGACCT 23

DB 314 ACTGCTGAATGTTGAGAGACCT 293

RESULT 11  
AAC46212/C

ID	AAC46212 standard; DNA; 963 BP.	PR	18-JUN-1999;	990S-0139461.
XX		PR	18-JUN-1999;	990S-0139462.
AC	AAC46212;	PR	18-JUN-1999;	990S-0139463.
XX		PR	18-JUN-1999;	990S-0139750.
DT	18-OCT-2000 (first entry)	PR	18-JUN-1999;	990S-0139763.
XX		PR	21-JUN-1999;	990S-0139817.
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49311.	PR	22-JUN-1999;	990S-0139899.
XX		PR	23-JUN-1999;	990S-0140353.
KW	Hybridisation assay; genetic mapping; gene expression control;	PR	23-JUN-1999;	990S-0140354.
KM	Protein identification; signal transduction pathway;	PR	24-JUN-1999;	990S-0140695.
KM	metabolic pathway; promoter; termination sequence; ss.	PR	28-JUN-1999;	990S-0140823.
XX		PR	29-JUN-1999;	990S-0140891.
OS	Arabidopsis thaliana.	PR	30-JUN-1999;	990S-0141287.
XX		PR	01-JUL-1999;	990S-0141842.
XX	EP1033405-A2.	PR	01-JUL-1999;	990S-0142154.
PD	06-SEP-2000.	PR	02-JUL-1999;	990S-0142055.
XX		PR	06-JUL-1999;	990S-0142390.
XX		PR	08-JUL-1999;	990S-0142803.
PF	25-FEB-2000; 2000EP-0301439.	PR	09-JUL-1999;	990S-0142920.
XX		PR	12-JUL-1999;	990S-0142977.
XX		PR	13-JUL-1999;	990S-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	990S-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	990S-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	990S-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	990S-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	990S-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	990S-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	990S-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	990S-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	990S-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	990S-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	990S-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	990S-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	990S-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	990S-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	990S-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	990S-0145088.
PR	04-MAY-1999;	PR	22-JUL-1999;	990S-0145087.
PR	05-MAY-1999;	PR	22-JUL-1999;	990S-0145089.
PR	06-MAY-1999;	PR	22-JUL-1999;	990S-0145192.
PR	07-MAY-1999;	PR	23-JUL-1999;	990S-0145145.
PR	11-MAY-1999;	PR	23-JUL-1999;	990S-0145218.
PR	14-MAY-1999;	PR	23-JUL-1999;	990S-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	990S-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	990S-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	990S-0145918.
PR	18-MAY-1999;	PR	27-JUL-1999;	990S-0145919.
PR	19-MAY-1999;	PR	28-JUL-1999;	990S-0145951.
PR	20-MAY-1999;	PR	02-AUG-1999;	990S-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	990S-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	990S-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	990S-0147038.
PR	27-MAY-1999;	PR	04-AUG-1999;	990S-0147204.
PR	28-MAY-1999;	PR	05-AUG-1999;	990S-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	990S-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	990S-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	990S-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	990S-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	990S-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	990S-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	990S-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	990S-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	990S-0148341.
PR	16-JUN-1999;	PR	13-AUG-1999;	990S-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	990S-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	990S-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	990S-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	990S-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	990S-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	990S-0149723.
PR	18-JUN-1999;	PR	23-AUG-1999;	990S-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	990S-0149902.

PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 963;  
Best Local Similarity 86.4%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGACC 22  
DB 89 GAGTGTGATGTTGAGAGACC 68  
RESULT 12  
AAC4514/c  
ID AAC4514 standard; DNA; 1054 BP.  
XX AAC4514;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 46771.

XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 03-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145189.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148171.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 20-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.

PR 31-AUG-1999; 990S-0151438.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155486.  
PR 24-SEP-1999; 990S-0155659.  
PR 28-SEP-1999; 990S-0156458.  
PR 29-SEP-1999; 990S-0156596.  
PR 04-OCT-1999; 990S-0157117.  
PR 05-OCT-1999; 990S-0157753.  
PR 06-OCT-1999; 990S-0157865.  
PR 07-OCT-1999; 990S-0158029.  
PR 08-OCT-1999; 990S-0158232.  
PR 12-OCT-1999; 990S-0158369.  
PR 13-OCT-1999; 990S-0159293.  
PR 13-OCT-1999; 990S-0159294.  
PR 13-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159329.  
PR 14-OCT-1999; 990S-0159330.  
PR 14-OCT-1999; 990S-0159331.  
PR 14-OCT-1999; 990S-0159637.  
PR 14-OCT-1999; 990S-0159638.  
PR 18-OCT-1999; 990S-0159584.  
PR 21-OCT-1999; 990S-0160741.  
PR 21-OCT-1999; 990S-0160767.  
PR 21-OCT-1999; 990S-0160768.  
PR 21-OCT-1999; 990S-0160770.  
PR 21-OCT-1999; 990S-0160814.  
PR 21-OCT-1999; 990S-0160815.  
PR 22-OCT-1999; 990S-0160980.  
PR 22-OCT-1999; 990S-0160981.  
PR 22-OCT-1999; 990S-0160989.  
PR 25-OCT-1999; 990S-0161404.  
PR 25-OCT-1999; 990S-0161405.  
PR 25-OCT-1999; 990S-0161406.  
PR 26-OCT-1999; 990S-0161359.  
PR 26-OCT-1999; 990S-0161359.  
PR 26-OCT-1999; 990S-0161360.  
PR 26-OCT-1999; 990S-0161361.  
PR 28-OCT-1999; 990S-0161920.  
PR 28-OCT-1999; 990S-0161992.  
PR 28-OCT-1999; 990S-0161993.  
PR 29-OCT-1999; 990S-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1054;  
Best Local Similarity 86.4%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGTGGTATGTCAGAGAGCC 22  
|||||||  
Db 152 GAGTGGTATGTCAGAGAGTAC 131

RESULT 13  
AAC37611/C  
ID AAC37611 standard; DNA; 1056 BP.

AC AAC37611;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18019.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN	EP103405-A2.	
PD		
XX	06-SEP-2000.	
XX		
PF	25-FEB-2000;	2000EE-0301439.
XX		
PR	05-MAR-1999;	9905-0121825
PR	05-MAR-1999;	9905-0123180.
PR	09-MAR-1999;	9905-0125348.
PR	23-MAR-1999;	9905-0125788.
PR	25-MAR-1999;	9905-0126564.
PR	29-MAR-1999;	9905-0126785.
PR	01-APR-1999;	9905-0127462.
PR	06-APR-1999;	9905-0128234.
PR	08-APR-1999;	9905-0128711.
PR	16-APR-1999;	9905-0129645.
PR	19-APR-1999;	9905-0130077.
PR	21-APR-1999;	9905-0130649.
PR	23-APR-1999;	9905-0130510.
PR	23-APR-1999;	9905-0130891.
PR	28-APR-1999;	9905-0131048.
PR	30-APR-1999;	9905-0132149.
PR	30-APR-1999;	9905-0132477.
PR	04-MAY-1999;	9905-0132484.
PR	05-MAY-1999;	9905-0132485.
PR	06-MAY-1999;	9905-0132486.
PR	06-MAY-1999;	9905-0132487.
PR	07-MAY-1999;	9905-0134768.
PR	11-MAY-1999;	9905-0134256.
PR	14-MAY-1999;	9905-0134218.
PR	14-MAY-1999;	9905-0134221.
PR	14-MAY-1999;	9905-0134421.
PR	14-MAY-1999;	9905-0134768.
PR	18-MAY-1999;	9905-0134768.
PR	19-MAY-1999;	9905-0134641.
PR	20-MAY-1999;	9905-0135124.
PR	21-MAY-1999;	9905-0135353.
PR	24-MAY-1999;	9905-0135629.
PR	25-MAY-1999;	9905-0136021.
PR	27-MAY-1999;	9905-0136592.
PR	28-MAY-1999;	9905-0136782.
PR	01-JUN-1999;	9905-0137722.
PR	03-JUN-1999;	9905-0137528.
PR	04-JUN-1999;	9905-0137502.
PR	07-JUN-1999;	9905-0137724.
PR	08-JUN-1999;	9905-0138094.
PR	10-JUN-1999;	9905-0138540.
PR	10-JUN-1999;	9905-0138647.
PR	14-JUN-1999;	9905-0138847.
PR	16-JUN-1999;	9905-0139452.
PR	16-JUN-1999;	9905-0139453.
PR	17-JUN-1999;	9905-0139492.
PR	18-JUN-1999;	9905-0139454.
PR	18-JUN-1999;	9905-0139455.
PR	18-JUN-1999;	9905-0139456.
PR	18-JUN-1999;	9905-0139457.
PR	18-JUN-1999;	9905-0139458.
PR	18-JUN-1999;	9905-0139459.
PR	18-JUN-1999;	9905-0139460.
PR	18-JUN-1999;	9905-0139461.
PR	18-JUN-1999;	9905-0139462.
PR	18-JUN-1999;	9905-0139463.
PR	18-JUN-1999;	9905-0139750.
PR	18-JUN-1999;	9905-0139763.
PR	21-JUN-1999;	9905-0139617.
PR	22-JUN-1999;	9905-0139699.
PR	23-JUN-1999;	9905-0140353.
PR	23-JUN-1999;	9905-0140354.
PR	24-JUN-1999;	9905-0140693.
PR	28-JUN-1999;	9905-0140823.
PR	29-JUN-1999;	9905-0140991.
PR	30-JUN-1999;	9905-0141847.
PR	01-JUL-1999;	9905-0141842.

PR	01-JUL-1999;	9905-0142154
PR	02-JUL-1999;	9905-0142055
PR	06-JUL-1999;	9905-0142390
PR	08-JUL-1999;	9905-0142803
PR	09-JUL-1999;	9905-0142820
PR	12-JUL-1999;	9905-0142877
PR	13-JUL-1999;	9905-0143344
PR	14-JUL-1999;	9905-0143624
PR	15-JUL-1999;	9905-0144085
PR	16-JUL-1999;	9905-0144085
PR	16-JUL-1999;	9905-0144325
PR	19-JUL-1999;	9905-0144331
PR	19-JUL-1999;	9905-0144332
PR	19-JUL-1999;	9905-0144332
PR	19-JUL-1999;	9905-0144334
PR	19-JUL-1999;	9905-0144334
PR	20-JUL-1999;	9905-0144352
PR	20-JUL-1999;	9905-0144352
PR	20-JUL-1999;	9905-0144352
PR	20-JUL-1999;	9905-0144352
PR	21-JUL-1999;	9905-0144518
PR	21-JUL-1999;	9905-0145086
PR	21-JUL-1999;	9905-0145085
PR	21-JUL-1999;	9905-0145085
PR	22-JUL-1999;	9905-0145087
PR	22-JUL-1999;	9905-0145089
PR	22-JUL-1999;	9905-0145192
PR	23-JUL-1999;	9905-0145218
PR	23-JUL-1999;	9905-0145218
PR	23-JUL-1999;	9905-0145224
PR	26-JUL-1999;	9905-0145276
PR	27-JUL-1999;	9905-0145313
PR	27-JUL-1999;	9905-0145318
PR	27-JUL-1999;	9905-0145519
PR	28-JUL-1999;	9905-0145551
PR	02-AUG-1999;	9905-0146386
PR	02-AUG-1999;	9905-0146388
PR	02-AUG-1999;	9905-0146389
PR	03-AUG-1999;	9905-0147038
PR	04-AUG-1999;	9905-0147204
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147192
PR	05-AUG-1999;	9905-0147260
PR	06-AUG-1999;	9905-0147303
PR	06-AUG-1999;	9905-0147416
PR	09-AUG-1999;	9905-0147493
PR	10-AUG-1999;	9905-0148171
PR	11-AUG-1999;	9905-0148341
PR	12-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148365
PR	13-AUG-1999;	9905-0148684
PR	16-AUG-1999;	9905-0149175
PR	17-AUG-1999;	9905-0149176
PR	18-AUG-1999;	9905-0149426
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149923
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151530
PR	07-SEP-1999;	9905-0152363
PR	10-SEP-1999;	9905-0153070
PR	13-SEP-1999;	9905-0153070
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039

PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158389.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1056;  
Best Local Similarity 86.4%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 1 GAGTGTGATGTGAGAGCC 22  
DB 154 GAGTGTGATGTGAGAGTGC 133

RESULT 14  
ID AAA79490  
AAA79490 standard; cDNA; 2884 BP.

AC AAA79490;  
DT 27-NOV-2000 (first entry)  
DE Eucalyptus grandis cell signalling involved polynucleotide SEQ ID NO:291.  
XX  
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
KW environmental change; development; cell proliferation; differentiation;  
KW elongation; survival; disease resistance; nutrient metabolism; ss.  
XX  
OS Eucalyptus grandis.  
XX  
PN WO200042171-A1.  
XX  
XX 20-JUL-2000.  
XX  
XX 11-JAN-2000; 2000MO-US00724.  
XX  
XX

PR 12-JAN-1999; 99US-0228986.  
PR 01-NOV-1999; 99US-0162866.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
PI Strabala TJ, Nieuwenhuizen NJ;  
XX  
XX WPI; 2000-476052/41.  
DR  
XX  
XX Isolated polynucleotide encoding a polypeptide involved in cell  
PT signaling used for generating transgenic plants with modified responses  
PT to external signals -  
XX  
PS Claim 1; Page 152-153; 527pp; English.  
XX  
CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
CC are involved in cell signalling. The polynucleotide and protein  
CC sequences can be used to modify the response of plant cells to external  
CC signals e.g. environmental changes or pathogens during the growth and  
CC development of a plant. They can be used to modify cell proliferation,  
CC differentiation, elongation and survival, resistance to disease and  
CC nutrient metabolism. Examples of modifications which can be produced are  
CC altered fruit ripening and senescence of leaves and flowers e.g. to  
CC delay senescence and prolong the life of cut flowers or enhance  
CC senescence of reproductive organs to engineer sterile plants. Other  
CC modifications can be used to delay senescence in selected cell types or  
CC organs providing fruit and vegetables which have a longer shelf life  
CC between harvest and consumption, or to decrease branching frequency in  
CC forest tree species giving long stretches of valuable knot-free clear  
CC wood which can be used in solid timber furniture and veneers.  
XX  
SQ Sequence 2884 BP; 856 A; 627 C; 705 G; 696 T; 0 other;

Query Match 74.8%; Score 17.2; DB 21; Length 2884;  
Best Local Similarity 86.4%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 2 AGTGTGATGTGAGAGCCT 23  
DB 2701 AGTGTGATGTGAGAGCCT 2722

RESULT 15  
ID AAA79657  
AAA79657 standard; cDNA; 3600 BP.

AC AAA79657;  
DT 27-NOV-2000 (first entry)  
DE Eucalyptus grandis cell signalling involved polynucleotide SEQ ID NO:458.  
XX  
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
KW environmental change; development; cell proliferation; differentiation;  
KW elongation; survival; disease resistance; nutrient metabolism; ss.  
XX  
OS Eucalyptus grandis.  
XX  
PN WO200042171-A1.  
XX  
XX 20-JUL-2000.  
XX  
XX 11-JAN-2000; 2000MO-US00724.  
XX  
XX 12-JAN-1999; 99US-0228986.  
XX  
XX 01-NOV-1999; 99US-0162866.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
XX Strabala TJ, Nieuwenhuizen NJ;  
XX  
XX

XX WPI: 2000-476052/41.  
DR  
XX

PT Isolated polynucleotide encoding a polypeptide involved in cell  
PT signaling used for generating transgenic plants with modified responses  
PT to external signals -  
XX  
XX

PS Claim 1; Page 203-204; 527pp; English.  
XX

CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
CC are involved in cell signalling. The polynucleotide and protein  
CC sequences can be used to modify the response of plant cells to external  
CC signals e.g. environmental changes or pathogens during the growth and  
CC development of a plant. They can be used to modify cell proliferation,  
CC differentiation, elongation and survival, resistance to disease and  
CC nutrient metabolism. Examples of modifications which can be produced are  
CC altered fruit ripening and senescence of leaves and flowers e.g. to  
CC delay senescence and prolong the life of cut flowers or enhance  
CC senescence of reproductive organs to engineer sterile plants. Other  
CC modifications can be used to delay senescence in selected cell types or  
CC organs providing fruit and vegetables which have a longer shelf life  
CC between harvest and consumption, or to decrease branching frequency in  
CC forest tree species giving long stretches of valuable knot-free clear  
CC wood which can be used in solid timber furniture and veneers.  
XX

SO Sequence 3600 BP; 1026 A; 811 C; 843 G; 920 T; 0 other;

Query Match

Best Local Similarity 74.8%; Score 17.2; DB 21; Length 3600;  
Matches 19; Conservativity 86.4%; Pred. No. 2.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 AGTGTGATGTTGAGAGAGCCT 23  
||||| |||||||||

DB 3417 AGTGGCCATGTTGAGAGAGCCTT 3438

Search completed: January 24, 2003, 19:45:32  
Job time : 6.27201 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:32 : Search time 56.8981 Seconds  
(without alignments)  
11764.270 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23

Sequence: 1 gagtggatgtgtgagagagcct 23

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hgt: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sy: \*  
12: gb\_un: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hgt\_hum: \*  
31: em\_hgt\_inv: \*  
32: em\_hgt\_other: \*  
33: em\_hgt\_mus: \*  
34: em\_hgt\_pln: \*  
35: em\_hgt\_rod: \*  
36: em\_hgt\_mam: \*  
37: em\_hgt\_vrt: \*  
38: em\_sy: \*  
39: em\_hgt\_hum: \*  
40: em\_hgt\_mus: \*  
41: em\_hgt\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	23	100.0	23	6 AX027825	AX027825 Sequence
2	23	100.0	3060	10 MUSVILLIN	M88454 Mus musculus
3	23	100.0	3104	10 BC015267	BC015267 Mus muscu
4	23	100.0	192060	2 AC098570	AC098570 Mus muscu
5	21.4	93.0	1501	9 BC017303	BC017303 Homo sapi
6	21.4	93.0	2702	6 A07400	A07400 H.sapiens D
7	21.4	93.0	2702	6 HSYILLR	X12901 Human mRNA
8	21.4	93.0	149109	2 AC101018	AC101018 Rattus no
9	21.4	93.0	149425	2 AC013320	AC013320 Homo sapi
10	21.4	93.0	159492	2 AC115173	AC115173 Rattus no
11	21.4	93.0	191754	9 AC021016	AC021016 Homo sapi
12	19.8	86.1	173085	2 AC112105	AC112105 Rattus no
13	18.8	81.7	18705	3 CEY266FA	AL023830 Caenorhab
14	18.8	81.7	86381	2 AL360077	AL360077 Homo sapi
15	18.8	81.7	100472	2 AP002875	AP002875 Homo sapi
16	18.8	81.7	131239	9 HS667H12	AL035414 Human DNA
17	18.8	81.7	143080	9 AC026431	AC026431 Homo sapi
18	18.8	81.7	144154	2 AP003762	AP003762 Oryza sat
19	18.8	81.7	152586	2 AC115073	AC115073 Mus muscu
20	18.8	81.7	160216	2 AC093951	AC093951 Rattus no
21	18.8	81.7	166143	2 AL356960	AL356960 Homo sapi
22	18.8	81.7	167284	9 AC112719	AC112719 Homo sapi
23	18.8	81.7	169263	2 AC112724	AC112724 Rattus no
24	18.8	81.7	174255	2 AC129726	AC129726 Rattus no
25	18.8	81.7	181297	2 AC020675	AC020675 Homo sapi
26	18.8	81.7	185144	2 AC093764	AC093764 Rattus no
27	18.8	81.7	185245	2 AC128046	AC128046 Rattus no
28	18.8	81.7	185932	2 AP003714	AP003714 Oryza sat
29	18.8	81.7	186787	9 AC010267	AC010267 Homo sapi
30	18.8	81.7	187415	2 AC110802	AC110802 Homo sapi
31	18.8	81.7	229537	2 AC127312	AC127312 Mus muscu
32	18.4	80.0	4151	14 H2U03488	U03488 Heliothis z
33	18.4	80.0	4286	6 AR062598	AR062598 Sequence
34	18.4	80.0	4286	6 AR071501	AR071501 Sequence
35	18.4	80.0	4286	6 E10704	E10704 DNA sequence
36	18.4	80.0	14708	3 AE002952	AE002952 Drosophila
37	18.4	80.0	75162	2 AC018480	AC018480 Drosophila
38	18.4	80.0	112755	10 AC083887	AC083887 Mus muscu
39	18.4	80.0	138851	10 AL645938	AL645938 Mouse DNA
40	18.4	80.0	167706	2 AC095302	AC095302 Rattus no
41	18.4	80.0	169341	2 AC107753	AC107753 Mus muscu
42	18.4	80.0	209572	2 AC099615	AC099615 Mus muscu
43	18.2	79.1	1663	6 AX467017	AX467017 Sequence
44	18.2	79.1	9223	8 AF435120	AF435120 Aspergill
45	18.2	79.1	121116	2 AC024661	AC024661 Homo sapi

#### ALIGNMENTS

RESULT 1  
AX027825  
LOCUS AX027825 23 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 2 from Patent WO0034492.  
ACCESSION AX027825  
VERSION AX027825.1 GI:10188666  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 23)  
AUTHORS Robine,S., Louvard,D., Pinto,D. and Jaissier,F.  
TITLE Regulatory sequences of the mouse villin gene - use in transgenesis  
JOURNAL Patent: WO 0034492-A 2 15-JUN-2000;  
ROBINE SYLVIE (FR) ; INST CURIE (FR) ; LOUWARD DANIEL (FR) ; PINTO

FEATURES	DANIEL (FR) ; CENTRE NAT RECH SCIENT (FR) ; JAISSER FREDERIC (FR)									
SOURCE	1. 23 Location/Qualifiers									
BASE COUNT	5 a 2 c 10 g 6 t									
ORIGIN										
Query Match	100.0%	Score 23;	DB 6;	length 23;						
Best Local Similarity	100.0%	Pred. No. 1.2;								
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;						
Oy	1	GAGTGTGATGTTGAGAGACCT	23							
Db	1	GAGTGTGATGTTGAGAGACCT	23							
RESULT 2										
LOCUS	MUSVILLIN	3660 bp	mRNA	linear	ROD 27-APR-1993					
DEFINITION	Mus musculus villin protein mRNA, complete cds.									
ACCESSION	M98454									
VERSION	M98454.1	GI:202365								
KEYWORDS	villin.									
SOURCE	Mus musculus (strain FVB) (library: lambda ZAP) adult intestinal epithelial cDNA to mRNA.									
ORGANISM	Mus musculus									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 3660)									
AUTHORS	Nit,D.A., Ezzell,R.M., Leung,J., Simlster,N.E., de Arruda,M.V. and Matsudaira,P.T.									
JOURNAL	Unpublished (1992)									
FEATURES	Location/Qualifiers									
source	1. 3060 /organism="Mus musculus" /strain="FVB" /db_xref="taxon:10090" /tissue_type="intestinal epithelial" /dev_stage="adult" /tissue_1fb="lambda ZAP" 55. .2538 /codon_start=1 /product="Villin" /protein_id="AAA0554.1" /db_xref="GI:202366" /translation="MTKLNQAVKSGSLNTTPGIIIMRIEAMQWVPVSPFSGFDDG CYVLAIIHTKSSITSLYDIHIIIGDSSDDEGAALITTDQDYLTKGRVDRPVCNQ ESSEIFRSYFQGLYIRKGGVAGASGKHHTNSCDVQRLHLHAKGRNVLAGEYEMSKFS NRGVFLFLIDGLKLIIONNGPESNMRELRCPILAKETRDGEGRTVGVVDEGEEDD SPQLMAIMNVLHGLRKLKLAISDSVAPAKALKLVHSDSGKLVRVAVATPPLT QDLKLHEDCYLIDQGLKLIKFWKGNKNAQERSAMQALNFIKAKQYPTQVEYVON VDGASPIFEOLFOKWTVPNRTSGKGTHTVSGVAKVDEKFDALTMHVQVAPADQKMK VDSGSEGVQWRIEDLELVPESKGLGHFGDCYLLIYTLIGERKHLYLIYIMQSGO ASODEIAASATQAVLLDOKYNDDEVQIIRVITMGKRPPIHMSITFKRMVYVTOGSTRKNIN LEPVPSTRFQVGRGTNADNTKAEVETARATSLNSNDVFLTKTPSCYLLMGCKGSGDB REKMAVADTISRTKEQVVGGEPAFNMWALGGKAPYANTKRLOEENOVITPPLFPCB SNQGRFLATEIFPDNDDLEEDVFLFDVDFEWMIGKHANEKKAAATTVQEVIL KTHGRNDLEETPIIVKOGHEPPTFGMFLAMDPEFKMSNTRSYDNLKAEIAGNSGDMSO IADVMSPKVDVFTANSTLSGRLPTPLPDLVAKSVEDLPQGVDPSPKREHLSIEDP TRALGMTPAAFSALPRKMQQNIKEKGLF"									
polya_signal	3039. .3045									
polya_site	3060									
BASE COUNT	770 a 794 c 849 g 647 t									
ORIGIN										
Query Match	100.0%	Score 23;	DB 10;	length 3060;						
Best Local Similarity	100.0%	Pred. No. 1.3;								
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;						
Oy	1	GAGTGTGATGTTGAGAGACCT	23							

[illegible]

ORIGIN

Query Match 100.0%; Score 23; DB 10; Length 3104;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGACCT 23  
 |||||||||||||||||||  
 DB 89 GAGTGTGATGTTGAGAGACCT 67

RESULT 4  
 AC098570/c 192060 bp DNA linear HMG 20-AUG-2002  
 LOCUS Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered  
 DEFINITION pieces.  
 AC098570  
 VERSION AC098570.2 GI:22325297  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULFILLTOP.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 192060)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP23-278N11  
 Unpublished  
 2 (bases 1 to 192060)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gaidyna,S., Glndu,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hages,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nobu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhkhong,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL Direct Submission  
 REFERENCE Submitted (24-OCT-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 AUTHORS 3 (bases 1 to 192060)  
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gaidyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hages,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhkhong,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnpbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
 JOURNAL Direct Submission

JOURNAL  
 COMMENT  
 Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 20, 2002 this sequence version replaced gi:16356898.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L14796  
 Center clone name: 278\_N\_11  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 190464 bases at least Q40  
 Consensus quality: 191061 bases at least Q30  
 Consensus quality: 191345 bases at least Q20  
 Insert size: 163000; agarose-1p  
 Insert size: 191560; sum-of-ctrls  
 Quality coverage: 12.4 in Q20 bases; agarose-1p  
 Quality coverage: 10.5 in Q20 bases; sum-of-ctrls  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 4921: contig of 4921 bp in length  
 \* 4922 5021: gap of 100 bp  
 \* 5022 15200: contig of 10179 bp in length  
 \* 15201 15300: gap of 100 bp  
 \* 15301 78692: contig of 63392 bp in length  
 \* 78693 78792: gap of 100 bp  
 \* 78793 116118: contig of 37326 bp in length  
 \* 116119 116218: gap of 100 bp  
 \* 116219 174491: contig of 58273 bp in length  
 \* 174492 174591: gap of 100 bp  
 \* 174592 192060: contig of 17469 bp in length.  
 Location/Qualifiers  
 1..192060  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RP23-278N11"  
 /clone\_lib="RP23-278N11"  
 /clone\_lib="RP23-278N11"  
 1..4921  
 /note="assembly-fragment  
 clone\_end:SP6  
 vector\_side:left"  
 5022..15200  
 /note="assembly-fragment"  
 15301..78692  
 /note="assembly-fragment"  
 78793..116118  
 /note="assembly-fragment"  
 116219..174491  
 /note="assembly-fragment"  
 174592..192060  
 /note="assembly-fragment  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 49668 a 46807 c 46516 g 48562 t 507 others

ORIGIN

Query Match 100.0%; Score 23; DB 2; Length 192060;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGAGCCT 23  
|||||  
Db 49339 GAGTGTGATGTTGAGAGAGCCT 49317

RESULT 5  
BC017303 1501 bp mRNA linear PRI 09-NOV-2001  
LOCUS Homo sapiens, similar to villin 1, clone MGC:29534 IMAGE:5087169,  
DEFINITION mRNA, complete cds.

ACCESSION BC017303  
VERSION BC017303.1 GI:16878196  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1501)  
AUTHORS Strausberg, R.  
TITLE Direct Submissions  
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,  
USA

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcaps-remail.nih.gov](mailto:gcaps-remail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Sequencing Center (NISC),  
Gaithersburg, Maryland:

Web site: [http://www.nisc.nih.gov/nisc\\_mgc.html](http://www.nisc.nih.gov/nisc_mgc.html)  
Contact: nisc\_mgc@nsl.nsl.gov  
Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stanlipop, S., Thomas, P.J.,  
Tongson, E.E., Touchman, J.W., Tsurgoun, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

FEATURES  
source  
1. 1501  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:29534 IMAGE:5087169"  
/tissue\_type="liver, hepatocellular carcinoma"  
/clone\_id="NH\_MGC\_100"  
/lab\_host="DH10B"  
/note="Vector: pOTB7"  
30..1295  
/codon\_start=-1  
/product="Similar to villin 1"  
/protein\_id="AAH17303.1"  
/db\_xref="GI:16878197"  
/translation="MTKLSAIVGSGSLNITPGIOWIRIAMQVVPVPSSTFGSPFD  
CYTILAIHKTRASSLSDIHYIKGODSSLDGGAAYITTMDFLGRAYOHREVOG  
ESSEAFRGYFQGLVIRKGVASGKHKHETNSYDORLLHVKGRNVAVGEWMSKSF  
NRDGVFLDLGKILIONNGPESTRMEERLAKELRDERGGRITVGVVDENELA  
SPKLEVMNHVILKRLKAAVDPVVEPALKALAHVSDSEGNLVNEVATRP  
ODLSHEDCYTIDOGGLKITYMKGKANEDEKAGSHALNFIKAKYPSSTVEYON  
DGAESAIFQOLFOKWTASNTSGIKHTYGSVGEAGAVREPSSMARATWSTT  
HPSLTCTIFEDFAGSLVADGDVDTL"

BASE COUNT 394 a 378 c 468 g 261 t  
ORIGIN

Query Match 93.0%; Score 21.4; DB 9; Length 1501;  
Best Local Similarity 95.7%; Pred. No. 7.8;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGAGCCT 23  
|||||  
Db 78 GAGTGTGATGTTGAGAGAGCCT 56

RESULT 6  
A07400/c 2702 bp DNA linear PAT 20-JUL-1993  
LOCUS H.sapiens DNA for villin.  
DEFINITION A07400  
ACCESSION A07400.1 GI:412171  
VERSION villin.  
KEYWORDS villin.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2702)  
AUTHORS Arpin, M., Pringault, E., Garcia, A. and Louvard, D.  
TITLE Specific nucleic-acid fragments of the human villin gene and their  
Application to diagnostic ends  
JOURNAL PASTEUR  
INSTITUT

FEATURES  
source  
1. 2702  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
25..2508  
/codon\_start=-1  
/product="villin"  
/protein\_id="CAA00664.1"  
/db\_xref="GI:412172"  
/db\_xref="SWISS-PROT:P09327"  
/translation="MTKLSAIVGSGSLNITPGIOWIRIAMQVVPVPSSTFGSPFD  
CYTILAIHKTRASSLSDIHYIKGODSSLDGGAAYITTMDFLGRAYOHREVOG  
ESSEAFRGYFQGLVIRKGVASGKHKHETNSYDORLLHVKGRNVAVGEWMSKSF  
NRDGVFLDLGKILIONNGPESTRMEERLAKELRDERGGRITVGVVDENELA  
SPKLEVMNHVILKRLKAAVDPVVEPALKALAHVSDSEGNLVNEVATRP  
ODLSHEDCYTIDOGGLKITYMKGKANEDEKAGSHALNFIKAKYPSSTVEYON  
DGAESAIFQOLFOKWTASNTSGIKHTYGSVGEAGAVREPSSMARATWSTT  
HPSLTCTIFEDFAGSLVADGDVDTL"

CDS

BASE COUNT 681 a 716 c 780 g 524 t 1 others  
ORIGIN  
Query Match 93.0%; Score 21.4; DB 6; Length 2702;  
Best Local Similarity 95.7%; Pred. No. 7.8;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGAGCCT 23  
|||||  
Db 73 GAGTGTGATGTTGAGAGAGCCT 51

RESULT 7  
HSVILR/c 2702 bp mRNA linear PRI 12-SEP-1993  
LOCUS Human mRNA for villin.  
DEFINITION X12901  
ACCESSION X12901.1 GI:37842  
KEYWORDS actin-binding protein; calcium binding protein; villin.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 2702)  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source  
 1. 2702  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="H929"  
 /tissue\_type="intestine"  
 25. 2508  
 /note="Villin (AA 1 - 827)"  
 /codon\_start=1  
 /protein\_id="CAA31386.1"  
 /db\_xref="GI:37843"  
 /db\_xref="SWISS-PROT:P09327"  
 /translation="MTKLSAOKGSLNITTPGDIOMRIEAMQWVPDSSTFGSFFDGD  
 CYLLAIHKTAASSLYDIHWYIGDSSIDEAGAAIYTTOMDDFLKRAVQHRVQGN  
 ESEAFRCYKQGLIRKGVASGKMHETNSYDQRLHVKRNNVAVGVYGGVDSNLA  
 NRGDVFLLDGLKLIIONNGPESTRERLAKETLAKEDRGRTVYGVYGGVDSNLA  
 SPRLMEVNNHVLKGRRLKAAVPTVPEVPAKALIKLYHSDSGNLYVRAVATRP  
 ODLSHEDCYTIDGCGIKIYVWKRKANEDEKRCAMSHALNFIKAKOYPTETVEON  
 DGASAVFOOLFKQWTSNRTSGLKHIVGSAKVEDVFKDATSMHKVPQVAAQQA  
 VDDSGEVQWRLENLVPDSSKWLGHFGGDEYLLLYLLIGEKOHYLLVWQGSQ  
 ASDSEITASAQAVILDOKNKEPVOIRVPMGKEPFLMSIFKGRVYVYOGSTRTN  
 LETPSTRLEFOVGTGANNKAFEPAPARNLNSNDVFLTOSCCYLCKGSGSDE  
 REMAKVADITSTREKOVYVGGEPANPMWALGKAPANTKRIQSENLVITPTLPEC  
 SNKTGRFLATPEIPDNDLLEDDVFLIDWQVFPVIGKIANEELKAAATTAQEL  
 KTHSGRDPETPIIVYKQGHPPFTFCFLAMDPPKSNKTSYEDLKAESGNLDMWSO  
 ITAEVTSPPKDVFNANSNLSSGPIPELEQLVNRKPVLEPEGVDPSPKREHLSIEDP  
 TQAFGMPAFSAALPRMKOONLKEKGLF"

BASE COUNT 681 a 716 c 781 g 524 t  
 ORIGIN

Query Match 93.0%; Score 21.4; DB 9; Length 2702;  
 Best Local Similarity 95.7%; Pred. No. 7.8;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GAGTGGTATGTGAGAGCCT 23  
 Db 73 GGGTGGTATGTGAGAGCCT 51

RESULT 8  
 AC101018/c  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone RP32-328P7, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 9 unordered pieces.  
 AC101018  
 AC101018.2 GI:22381320  
 HTG: HTGS.PHASE1; HTGS\_FUULLTOP; HTGS\_ACTIVEFIN.  
 KEYWORDS  
 Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 149109)  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 REFERENCE  
 AUTHORS  
 Birren, B., Nussbaum, C., and Lander, E.  
 Rattus norvegicus, clone RP32-328P7  
 Unpublished  
 2 (bases 1 to 149109)  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Glnde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,  
 Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, D., Levine, R., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McNeeters, R., Meldrim, J.,  
 Meneus, L., Mihova, T., Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.  
 Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 149109)  
 REFERENCE  
 JOURNAL  
 TITLE  
 DIRECT SUBMISSION  
 AUTHORS  
 Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Gardyna, S., Gord, S., Graham, L., Grand-pierre, N., Hagos, B.,  
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,  
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,  
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, Y.,  
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A., and Zody, M.  
 Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced gi:17059792.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L14794  
 Center clone name: 328\_P-7

NOTE: This is a 'working draft' sequence. It currently  
 consists of 9 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 46829: contig of 46829 bp in length  
 \* 46830 46928: gap of 100 bp  
 \* 46930 49621: contig of 2692 bp in length  
 \* 49622 49721: gap of 100 bp  
 \* 49722 70009: contig of 20288 bp in length  
 \* 70010 70109: gap of 100 bp





2396 2495: gap of unknown length  
2496 3600: contig of 1105 bp in length  
2497 3700: gap of unknown length  
3701 5021: contig of 1321 bp in length  
5022 5121: gap of unknown length  
5122 6727: contig of 1606 bp in length  
6728 6828: gap of unknown length  
6828 8608: contig of 1781 bp in length  
8609 8709: gap of unknown length  
8709 10004: contig of 1295 bp in length  
10004 10103: gap of unknown length  
10104 11211: contig of 1108 bp in length  
11212 11311: gap of unknown length  
11312 12518: contig of 1207 bp in length  
12519 12618: gap of unknown length  
12619 13843: contig of 1225 bp in length  
13844 13943: gap of unknown length  
13944 15404: contig of 1461 bp in length  
15405 15504: gap of unknown length  
15505 17647: contig of 2143 bp in length  
17648 17747: gap of unknown length  
17748 19706: contig of 1959 bp in length  
19707 19806: gap of unknown length  
19807 21576: contig of 1770 bp in length  
21577 21676: gap of unknown length  
21677 22906: contig of 1230 bp in length  
22907 23006: gap of unknown length  
23007 25086: contig of 2080 bp in length  
25087 25186: gap of unknown length  
25187 26832: contig of 1646 bp in length  
26833 26932: gap of unknown length  
26933 28710: contig of 1778 bp in length  
28711 28810: gap of unknown length  
28811 30606: contig of 1796 bp in length  
30607 30706: gap of unknown length  
30707 33113: contig of 2407 bp in length  
33114 33213: gap of unknown length  
33214 35361: contig of 2148 bp in length  
35362 35461: gap of unknown length  
35462 37520: contig of 2059 bp in length  
37521 37620: gap of unknown length  
37621 40471: contig of 2851 bp in length  
40472 40571: gap of unknown length  
40572 42584: contig of 2013 bp in length  
42585 42684: gap of unknown length  
42685 45547: contig of 2863 bp in length  
45548 45647: gap of unknown length  
45648 47838: contig of 2191 bp in length  
47839 47938: gap of unknown length  
47939 51229: contig of 3291 bp in length  
51230 51329: gap of unknown length  
51330 54005: contig of 2676 bp in length  
54006 54105: gap of unknown length  
54106 56683: contig of 2578 bp in length  
56684 56783: gap of unknown length  
56784 59361: contig of 2578 bp in length  
59362 59461: gap of unknown length  
59462 61588: contig of 2127 bp in length  
61589 61688: gap of unknown length  
61689 65712: contig of 4024 bp in length  
65713 65812: gap of unknown length  
65813 69339: contig of 3527 bp in length  
69340 69439: gap of unknown length  
69440 72570: contig of 3131 bp in length  
72571 72670: gap of unknown length  
72671 76896: contig of 4228 bp in length  
76897 76996: gap of unknown length  
76999 80760: contig of 3762 bp in length  
80761 80860: gap of unknown length  
80861 86313: contig of 5453 bp in length  
86314 86413: gap of unknown length  
86414 92231: contig of 5818 bp in length  
92232 92331: gap of unknown length

92332 99692: contig of 7361 bp in length  
99693 99792: gap of unknown length  
99793 106178: contig of 6386 bp in length  
106179 106278: gap of unknown length  
106279 114304: contig of 8026 bp in length  
114305 114404: gap of unknown length  
114405 122112: contig of 7708 bp in length  
122113 122212: gap of unknown length  
122213 129642: contig of 7430 bp in length  
129643 129742: gap of unknown length  
129743 138684: contig of 8942 bp in length  
138685 138784: gap of unknown length  
138785 149662: contig of 10878 bp in length  
149663 149762: gap of unknown length  
149763 159492: contig of 9730 bp in length.  
location/Qualifiers  
1.159492  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-286017"  
BASE COUNT 39083 a 36447 c 36055 g 40464 t 7443 others  
ORIGIN  
Query Match 93.0%; Score 21.4; DB 2; Length 159492;  
Best Local Similarity 95.7%; Pred. No. 8;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GAGTGGTGGTGGAGAGCCT 23  
Db 13423 GAGTGGTGGTGGAGAGCCT 13445  
RESULT 11  
AC021016/c 191754 bp DNA linear PRI 09-MAY-2001  
LOCUS AC021016 Homo sapiens BAC clone RP11-378A13 from 2, complete sequence.  
DEFINITION AC021016  
ACCESSION AC021016.4 GI:11120952  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 191754)  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 191754)  
AUTHORS Cordes,M., Maupin,R., Hawkins,M. and Boyer,E.  
TITLE The sequence of Homo sapiens BAC clone RP11-378A13  
JOURNAL Unpublished  
3 (bases 1 to 191754)  
AUTHORS Waterston,R.H.  
TITLE Direct Submision  
JOURNAL Submitted (12-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 191754)  
AUTHORS Waterston,R.  
TITLE Direct Submision  
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington  
University 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 191754)  
AUTHORS Waterston,R.  
TITLE Direct Submision  
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 8, 2000 this sequence version replaced gl:7630969.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC



Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics -----  
Center project name: H\_NH0378A13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatematsu, M., Cataneese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

#### VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-146N10; the clone sequenced to the right is RP11-36003. Actual start of this clone is at base position 1 of RP11-378A13; actual end is at base position 191754 of RP11-378A13.

The sequence RP11-378A13 from 156426 to 156433 is derived from a single m13 subclone. Size of the region was confirmed by PCR from BAC DNA.

FEATURES  
Source  
1. Location/Qualifiers  
1. 191754  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-378A13"  
/clone\_id="RPCI-11"  
1. 224  
/note="similar to EST AA947005 (NID:g3110400) oq45e10.s1"  
1. 224  
/note="similar to EST A1184556 (NID:g3735194) qd60a10.x1"  
1. 224  
/note="similar to EST A1457236 (NID:g4310105) tk01d06.x1"  
1. 216  
/note="similar to EST AA782751 (NID:g2842082) aj08c02.s1"  
1. 216  
/note="similar to EST AW020425 (NID:g5873955) df09e11.y1"  
2. 223  
/note="similar to EST AA175052 (NID:g1756173) ms62f02.r1"  
2. 223  
/note="similar to EST AV079022 (NID:g5198850) "  
2. 223  
/note="similar to EST AW538472 (NID:g7180889) "  
2. 49  
/note="similar to EST AV330904 (NID:g6370956) "  
7. 224  
/note="similar to EST A1454294 (NID:g4292719) "  
7. 224  
misc\_feature

misc\_feature  
12. 223  
/note="similar to EST AW433991 (NID:g6365298) "  
20. 223  
/note="similar to EST AW059787 (NID:g6652109) "  
60. 220  
/note="similar to EST A1709077 (NID:g4998853) as86d11.x1"  
67. 223  
/note="similar to EST AA870364 (NID:g2865809) vq44e05.r1"  
447. 660  
/note="similar to EST AV330904 (NID:g6370956) "  
repeat\_region  
/rpt\_family="Alu"  
1182. 1425  
/rpt\_family="L1"  
1442. 1583  
/rpt\_family="MERL-type"  
1820. 1928  
/rpt\_family="Alu"  
1948. 2347  
/note="similar to EST AA224009 (NID:g1844621) zr12g06.r1"  
2162. 2293  
/rpt\_family="Alu"  
2297. 2587  
/rpt\_family="Alu"  
2597. 2804  
/rpt\_family="Alu"  
2606. 2923  
/note="similar to EST AW351965 (NID:g6849698) "  
2630. 3135  
/note="similar to EST AA234699 (NID:g1859192) zs39g05.r1"  
2804. 3242  
/note="similar to EST A1916611 (NID:g5636466) wa28f08.x1"  
2976. 3245  
/note="similar to EST A1401669 (NID:g4244756) th24e10.x1"  
3097. 3284  
/rpt\_family="MIR"  
3546. 3781  
/note="similar to EST AA102341 (NID:g1647297) zm26h05.r1"  
3848. 4127  
/rpt\_family="Alu"  
4187. 4469  
/rpt\_family="Alu"  
4474. 4690  
/rpt\_family="Alu"  
4723. 5241  
/note="similar to EST AW444974 (NID:g6386736) "  
4931. 5413  
/note="similar to EST AA100813 (NID:g1647248) zm26h05.s1"  
6006. 6137  
/rpt\_family="MIR"  
6243. 6394  
/rpt\_family="MIR"  
6967. 7092  
/note="similar to EST AW461516 (NID:g7031684) "  
7935. 8054  
/note="similar to EST AW461516 (NID:g7031684) "  
8294. 8433  
/rpt\_family="MERL-type"  
8553. 8604  
/note="similar to EST A1391439 (NID:g4217443) tf96e02.x1"  
8556. 8749  
/note="similar to EST AW461516 (NID:g7031684) "  
9362. 9733  
/note="similar to EST A1391439 (NID:g4217443) tf96e02.x1"  
9476. 9723  
/note="similar to EST AA595679 (NID:g2411029) nk66f01.s1"  
9997. 10265  
/note="similar to EST A1634626 (NID:g4685956) tz30g01.x1"  
10003. 10482  
/note="similar to EST A1828100 (NID:g5448771) wk31g06.x1"  
10003. 10482  
/note="similar to EST A1890574 (NID:g5595738) wm87e04.x1"  
10003. 10482  
/note="similar to EST A1978960 (NID:g5803990) wr8e06.x1"

```

msc_feature 10003..10482 /note="similar to EST AL047914 (NID:g4728747)"
msc_feature 10003..10482 /note="similar to EST AM027349 (NID:g5886105) w73e09.x1"
msc_feature 10003..10457 /note="similar to EST A1972200 (NID:g5769026) w76301.x1"
msc_feature 10003..10421 /note="similar to EST W48859 (NID:g1336988) zc42c08.s1"
msc_feature 10003..10330 /note="similar to EST AM074215 (NID:g6029213) xb09612.x1"
msc_feature 10003..10249 /note="similar to EST AA143563 (NID:g1712935) zc32310.s1"
msc_feature 10003..10247 /note="similar to EST A1583619 (NID:g4569516) ts16d09.x1"
msc_feature 10003..10240 /note="similar to EST A1583619 (NID:g4569516) ts16d09.x1"

Query Match 93.0% Score 21.4; DB 9; Length 191754;
Best Local Similarity 95.7% Pred. No. 8;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGGATGATGTGAGAGACCT 23
Db 169701 GGGTGGATGATGTGAGAGACCT 169679

RESULT 12
AC112105 173085 bp DNA linear HTG 17-JUL-2002
LOCUS Rattus norvegicus clone CH230-301N9, *** SEQUENCING IN PROGRESS
DEFINITION *** 58 unordered pieces.
ACCESSION AC112105
VERSION AC112105.3 GI:21744618
KEYWORDS HTG; HTGS.PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 173085)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,J., Benton,D., Blinze,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burke,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulky,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolyet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kretlow,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Maxwell,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabadi,K., Morgan,M., Morris,S.,
Moser,M., Nickerson,E., Nwokenwo,S., Ogum,M., Okunolu,G.,
Ogunyemi,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoan,I., Rolle,M., Ruiz,S., Savely,G.,
Scheer,S., Scott,G., Shen,H., Shoshari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

```

```

TANSEY,J., TAYLOR,C., TAYLOR,T., TELFORD,B., THOMAS,N., THOMAS,S.,
USMANI,K., VASQUEZ,L., VERA,V., VILLALON,D., VINSON,R., WANG,Q.,
WANG,S., WARD-MOORE,S., WARREN,R., WASHINGTON,C., WATLINGTON,S.,
WILLIAMS,G., WILLIAMSON,A., WLECZK,R., WOODEN,S., WORLEY,K.,
WU,C., WU,Y., WU,Y.F., ZHOU,J., ZOTTILLA,S., NELSON,D.,
WEINSTOCK,G., and GIBBS,R.
Direct Submission
Unpublished
2 (bases 1 to 173085)
Worley,K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173085)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:20303309.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSIW
Center clone name: CH230-301N9
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 123259 bases at least Q40
Consensus quality: 129338 bases at least Q30
Consensus quality: 134476 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1043 1042: contig of 1042 bp in length
1143 2142: contig of 1000 bp in length
2143 2242: gap of unknown length
2243 3440: contig of 1198 bp in length
3441 3540: gap of unknown length
3541 4828: contig of 1288 bp in length
4829 4928: gap of unknown length
4929 6007: contig of 1079 bp in length
6008 6107: gap of unknown length
6108 7456: contig of 1349 bp in length
7457 7556: gap of unknown length
7557 8803: contig of 1247 bp in length
8804 8904: gap of unknown length
8904 10250: contig of 1347 bp in length
10251 10350: gap of unknown length
10351 11457: contig of 1107 bp in length
11458 11557: gap of unknown length
11558 12846: contig of 1289 bp in length
12847 12946: gap of unknown length
12947 14228: contig of 1282 bp in length
14229 14328: gap of unknown length
14329 15874: contig of 1546 bp in length
15875 15974: gap of unknown length
15975 17615: contig of 1641 bp in length
17616 17715: gap of unknown length

```

```

* 17716 19098: contig of 1383 bp in length
* 19099 19198: gap of unknown length
* 19199 21275: contig of 2077 bp in length
* 21276 21375: gap of unknown length
* 21376 22377: contig of 1002 bp in length
* 22378 22477: gap of unknown length
* 22478 23752: contig of 1275 bp in length
* 23753 23853: gap of unknown length
* 23853 25173: contig of 1321 bp in length
* 25174 25273: gap of unknown length
* 25274 26931: contig of 1658 bp in length
* 26932 27031: gap of unknown length
* 27032 28858: contig of 1827 bp in length
* 28859 30192: gap of unknown length
* 30193 30292: contig of 1234 bp in length
* 30293 32429: contig of 2137 bp in length
* 32430 32529: gap of unknown length
* 32530 34037: contig of 1508 bp in length
* 34038 34137: gap of unknown length
* 34138 36540: contig of 2403 bp in length
* 36541 36641: gap of unknown length
* 36641 38123: contig of 1483 bp in length
* 38124 38224: gap of unknown length
* 38224 39986: contig of 1763 bp in length
* 39987 40087: gap of unknown length
* 40087 41898: contig of 1811 bp in length
* 41898 41997: gap of unknown length
* 41998 44140: contig of 2143 bp in length
* 44141 44240: gap of unknown length
* 44241 45410: contig of 1170 bp in length
* 45411 45510: gap of unknown length
* 45511 47629: contig of 2119 bp in length
* 47630 47729: gap of unknown length
* 47730 50367: contig of 2638 bp in length
* 50368 50467: gap of unknown length
* 50468 52623: contig of 2156 bp in length
* 52624 52723: gap of unknown length
* 52724 54648: contig of 1925 bp in length
* 54649 54748: gap of unknown length
* 54749 57379: contig of 2631 bp in length
* 57380 57479: gap of unknown length
* 57480 60063: contig of 2584 bp in length
* 60064 60163: gap of unknown length
* 60164 62589: contig of 2426 bp in length
* 62590 62689: gap of unknown length
* 62690 64485: contig of 1796 bp in length
* 64486 64585: gap of unknown length
* 64586 66646: contig of 2061 bp in length
* 66647 66746: gap of unknown length
* 66747 69761: contig of 3014 bp in length
* 69761 69860: gap of unknown length
* 69861 73271: contig of 3411 bp in length
* 73272 73372: gap of unknown length
* 73372 78287: contig of 4916 bp in length
* 78288 78387: gap of unknown length
* 78388 82141: contig of 3754 bp in length
* 82142 82241: gap of unknown length
* 82242 86068: contig of 3827 bp in length
* 86069 86168: gap of unknown length
* 86169 90383: contig of 4215 bp in length
* 90384 90483: gap of unknown length
* 90484 94073: contig of 3590 bp in length
* 94074 94173: gap of unknown length
* 94174 98119: contig of 3946 bp in length
* 98120 98220: gap of unknown length
* 98220 102403: contig of 4184 bp in length
* 102404 106962: gap of unknown length
* 106963 107062: contig of 4459 bp in length
* 107063 111515: gap of unknown length
* 111516 115636: contig of 4021 bp in length

```

```

* 115637 115736: gap of unknown length
* 115737 119071: contig of 3335 bp in length
* 119072 119171: gap of unknown length
* 119172 123459: contig of 4288 bp in length
* 123460 123559: gap of unknown length
* 123560 129565: contig of 6006 bp in length
* 129566 129666: gap of unknown length
* 129666 135079: contig of 5414 bp in length

Query Match      86.1%; Score 19.8; DB 2; Length 173085;
Best Local Similarity 91.3%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GAGTGTGATGTTGAGAGCCT 23
Db 102276 GAGTGTGTTGTGAGAGTCT 102298

RESULT 13
CEY26E6A/C
LOCUS      CEY26E6A      18705 bp      DNA      linear      INV 24-JAN-2002
DEFINITION Caenorhabditis elegans cosmid Y26E6A, complete sequence.
ACCESSION  AL023830 AL008874
VERSION     AL023830.1 GI:3217845
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
            Rhabdilitoidea; Rhabdilitidae; Peloderiinae; Caenorhabditis.

REFERENCE   1
            none.
            Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
            Science 282 (5396), 2012-2018 (1998)
            99069613
            PUBMED
            9851916
            The C. elegans Sequencing Consortium.
            2 (bases 1 to 18705)
            Lennard, N.
            Direct Submission
            Submitted (09-JUN-1998) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            joesanger.ac.uk or rwhematode.wustl.edu
            On May 14, 2001 this sequence version replaced gi:2612776.
            Coding sequences below are predicted from computer analysis, using
            predictions from GeneFINDER (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone Y26E6A.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone K02B9 is at 18602 in this sequence. The
            true right end of clone R21H8 is at 104 in this sequence. The start
            of this sequence (1..104) overlaps with the end of sequence 278546.
            The end of this sequence (18602..18705) overlaps with the start of
            sequence 269663.
            For a graphical representation of this sequence and its analysis
            see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=Y26E6A
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            Location/Qualifiers
                1..18705
                /organism="Caenorhabditis elegans"
                /db_xref="taxon:6239"

```

## FEATURES

## SOURCE

```
gene
/clone="y26E6A"
join(complement(1130..1285),complement(910..1064),
complement(737..863),complement(278546..1:37732..37919),
complement(278546..1:37485..37686),
complement(278546..1:37082..37381),
complement(72188..4"
join(complement(1130..1285),complement(910..1064),
complement(737..863),complement(278546..1:37732..37919),
complement(278546..1:37485..37686),
complement(278546..1:37082..37381))
/clone="T21H8.4"
/clone="similar to 7Tm receptor like"
/codon_start=1
/protein_id="CA19474.1"
/db_xref="GI:3880747"
/translation="MECETRGQLYHPAVLAIIFGTSGYITPILITLKIVAK
VYHPNIVFLMLNVSCLLGLTAMSAIPLNMIKPSDDLKTYFCISKIGT
FLFAPCLVITTSAGILERSMATYVKNYEROGALGTLIAVAVAATATILFLP
EDGEEMITTCITFASKISGRVYVMEFOLLDAVISIHLVLYRNKIDKSGTS
LSQFQRENWVKTKQVTPLLISNTYIGYITVSPFLRYKLYLPNNYEIIAALF
IMHPMPFSLITVELMIGKKRODIHMDMASODPDDOPIHMEMNDVIFQSL
KAQARANKRPLSANMTSTLITKFKKQTTVAASIPSPS"
complement(join(9082..9405,9771..10229,10368..10826))
/gene="y26E6A.2"
complement(join(9082..9405,9771..10229,10368..10826))
/gene="y26E6A.2"
/codon_start=1
/protein_id="CA19473.1"
/db_xref="GI:3880746"
/translation="MADLDLELFRVAPFPPPOAKDRVNNENMKEDMELLTWTVE
EISINGPTPRYRCLEIOLGETFEKSSALRGVGFETLDAIHCDFMAHLELT
KMWPSHLISRDNEFRPKSKNVTKHIIBDDYKLPDLKIPMITQIMORSYKKAQ
SLAITSAAANITRONSLSKVPDLVILIDFSSGEMNVRKNOQENSDSKK
LAIQEVNKKKQTLRLNTYILKIVTDDPECEAVHNHVSGRFKRIIVKQFTE
NLKMWEMEKVEYEVNALTDDIPFPIKISIRIDLRDGHGRKLSAKHFMV
CALNIDITFRKALRDMDNGHREVGRAWILKQFIQFPGRRHRPKTSPFKGLFETS
ELNNSOMPGEITEYEVHETN"
complement(join(13206..13571,13624..13776,13855..14061,
14129..14229,14280..14528,14753..15026))
/gene="y26E6A.1"
complement(join(13206..13571,13624..13776,13855..14061,
14129..14229,14280..14528,14753..15026))
/gene="y26E6A.1"
/clone="CDNA EST yk229h9.3 comes from this gene
CDNA EST yk458f3.3 comes from this gene
CDNA EST yk49794.3 comes from this gene
CDNA EST yk229h9.5 comes from this gene
CDNA EST yk458f3.5 comes from this gene
CDNA EST yk49794.5 comes from this gene
CDNA EST yk516f7.5 comes from this gene
CDNA EST yk567h10.5 comes from this gene"
/codon_start=1
/protein_id="CA19472.1"
/db_xref="GI:3880745"
/translation="MLPVANPEEQANBEELRQADIDPOARNHNRNARRRREME
RANRNRRLQADRGQAARERGPAGRAQARARARALRANNAQALQPDAAVLEA
LDEINIDFDEDEPNRRRDVQPOVLEFANLPRLIALDPDAQLANOLAMAHML
APENYIRVARERQSGFCIGKSSVHNAFLNHDTCASGACGKIMDMENHFLS
STYSLRQOLYKLEKRELEFEFARVNDPERLEHMEGACITCAHKLGNPNORVY
OIHANKIEKTVANHPFOFLKTEKEDGISTKTAHPDSIOCCIDGDPDEFQNR
HGTWYERKDTVEEQLEKRYLEETIEKILVYRSHREVTYLTILPMDKLACIEE
LKNSPRIARIQIRVHCIDPIRGVPLEGEAELEDDEDINAAYQAAEEFQLGIT"
```

```
BASE COUNT 6444 a 2981 c 3321 g 5959 t
ORIGIN
```

```
Query Match 81.7%; Score 18.8; DB 3; Length 18705;
Best Local Similarity 90.98; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 GAGTGTGATGTTGAGAGACC 22
Db 7751 GATTGGTGGATGTTGAGAGCC 7730
AL360077/c
LOCUS
DEFINITION
RESULT 14
AL360077/c
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 86381)
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humqueres@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Nov 23, 2000 this sequence version replaced gi:3801102.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humqueres@sanger.ac.uk
----- Project Information
Center project name: BK501D19
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Assembly vector: plasmid: 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 81882 bases at least Q40
Consensus quality: 83656 bases at least Q30
Consensus quality: 84751 bases at least Q20
Insert size: 85681; sum-of-contigs
Insert size: 102977; 7.3% error; agarose-fp
Quality coverage: 4.03x in Q20 bases; sum-of-contigs Quality
Coverage: 3.60x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 12686: contig of 12686 bp in length
12687 12786: gap of 100 bp
12787 16897: contig of 4111 bp in length
16898 16997: gap of 100 bp
16998 30118: contig of 13121 bp in length
30119 30218: gap of 100 bp
30219 33053: contig of 2835 bp in length
33054 33153: gap of 100 bp
33154 52695: contig of 19542 bp in length
52696 52795: gap of 100 bp
52796 59088: contig of 6293 bp in length
59089 59188: gap of 100 bp
59189 73338: contig of 14150 bp in length
73339 73438: gap of 100 bp
73439 86381: contig of 12943 bp in length.
Location/Qualifiers
1. 86381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTA-501D19"
/clone_lib="CIT978SK-A2"
```

misc\_feature 1..12686  
/note="assembly\_fragment:00068"  
misc\_feature 12787..16897  
/note="assembly\_fragment:00338"  
misc\_feature 16998..30118  
/note="assembly\_fragment:00394"  
misc\_feature 30219..33053  
/note="assembly\_fragment:00465"  
misc\_feature 33154..52695  
/note="assembly\_fragment:00486"  
misc\_feature 52796..59088  
/note="assembly\_fragment:00701"  
misc\_feature 59189..73338  
/note="assembly\_fragment:00876"  
misc\_feature 73439..86381  
/note="assembly\_fragment:00891"  
BASE COUNT 24179 a 19215 c 17718 g 24568 t 701 others  
ORIGIN

Query Match 81.7% Score 18.8; DB 2; length 86381;  
Best Local Similarity 90.9%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AGTGGTGAATGTTGAGAGACCT 23  
|||||  
DB 45356 AGTGGTGAATGTTGAGAGACCT 45335

RESULT 15  
AP002875 100472 bp DNA linear HTG 07-OCT-2000  
LOCUS Homo sapiens chromosome 4 clone 390M07 map 4q16-q20, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 2 ordered pieces.  
ACCESSION AP002875  
VERSION AP002875.1 GI:10716811  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Homo sapiens DNA, clone:390M07.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 100472)  
AUTHORS Tsai,S.F.  
TITLE Direct Submission  
JOURNAL Submitted (06-OCT-2000) Shih-Peng Tsai, National Yang-Ming University, Institute of Genetics, 155 Li-Rong St. Section 2, Peitou, Taipei, Taiwan 11221, Republic of China  
(E-mail:ympe@ym.edu.tw, URL:http://genome.ym.edu.tw/,  
Tel:886-2-28267043, Fax:886-2-28264930)

## COMMENT

\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 50663: contig of 50663 bp in length  
\* 50764 100472: contig of 49709 bp in length.  
Location/Qualifiers

## FEATURES

source 1..100472  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="4q16-q20"  
/clone="390M07"  
BASE COUNT 30972 a 19378 c 19589 g 30429 t 104 others  
ORIGIN

Query Match 81.7% Score 18.8; DB 2; length 100472;  
Best Local Similarity 90.9%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGTGGTGAATGTTGAGAGACCT 23  
|||||  
DB 8050 AGTGGTGAATGTTGAGAGACCT 8071

Search completed: January 25, 2003, 04:29:33  
Job time : 160.898 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:32 ; Search time 2252.1 Seconds  
(without alignments)  
11764.270 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctggtgaccagagaca.....ctctaggtctgcacacatg 8995

Scoring table:

IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rtd: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8995	100.0	8995	6 AX027824	AX027824 Sequence
2	8762.2	97.4	192060	2 AC098570	AC098570 Mus muscu
3	3016.2	33.5	149109	2 AC101018	AC101018 Rattus no
4	938.6	10.4	159492	2 AC115173	AC115173 Rattus no
5	742.6	8.3	159492	2 AC115173	AC115173 Rattus no
6	188.6	2.1	80578	2 AC119514	AC119514 Rattus no
7	188.6	2.1	181772	2 AC094169	AC094169 Rattus no
8	188.6	2.1	194630	2 AC128412	AC128412 Rattus no
9	188	2.1	149425	2 AC013320	AC013320 Homo sapi
10	184.8	2.1	191754	9 AC021016	AC021016 Homo sapi
11	184.8	2.0	220909	2 AC115965	AC115965 Mus muscu
12	184	2.0	184591	2 AC102777	AC102777 Mus muscu
13	183.6	2.0	273800	10 AF100956	AF100956 Mus muscu
14	181.8	2.0	182695	2 AC130278	AC130278 Mus muscu
15	179.2	2.0	218502	2 AC025964	AC025964 Mus muscu
16	179.2	2.0	240536	10 AC025910	AC025910 Mus muscu
17	179.2	2.0	241400	10 AL646093	AL646093 Mouse DNA
18	178.2	2.0	198923	2 AC109263	AC109263 Mus muscu
19	177.8	2.0	165345	2 AC117095	AC117095 Rattus no
20	177.8	2.0	179805	2 AC122956	AC122956 Rattus no
21	177.6	2.0	224389	2 AC123811	AC123811 Mus muscu
22	177.4	2.0	208161	2 AC074145	AC074145 Mus muscu
23	177.2	2.0	165259	2 AL844206	AL844206 Mus muscu
24	177.2	2.0	166167	10 AL611936	AL611936 Mouse DNA
25	177	2.0	215049	2 AC122807	AC122807 Mus muscu
26	176.6	2.0	204573	10 AC087541	AC087541 Mus muscu
27	175.8	2.0	200057	2 AC073810	AC073810 Mus muscu
28	175	1.9	88410	2 AC098985	AC098985 Rattus no
29	174.4	1.9	207814	10 AL590994	AL590994 Mouse DNA
30	174.2	1.9	202095	2 AC123423	AC123423 Mus muscu
31	174	1.9	168280	2 AC123357	AC123357 Rattus no
32	173.6	1.9	119657	2 AC073667	AC073667 Mus muscu
33	173.6	1.9	170139	2 AC073727	AC073727 Mus muscu
34	172.8	1.9	249283	2 AC124577	AC124577 Mus muscu
35	172.4	1.9	73838	10 AL646020	AL646020 Mouse DNA
36	171.4	1.9	220242	10 AL603706	AL603706 Mouse DNA
37	171.2	1.9	186127	2 AC102630	AC102630 Mus muscu
38	170.6	1.9	159169	2 AC113976	AC113976 Mus muscu
39	170.4	1.9	188168	2 AC115806	AC115806 Mus muscu
40	170.4	1.9	195430	2 AC116140	AC116140 Mus muscu
41	170.4	1.9	207696	2 AC093473	AC093473 Mus muscu
42	170.2	1.9	135267	2 AC021434	AC021434 Mus muscu
43	170.2	1.9	158338	2 AL845258	AL845258 Mus muscu
44	170	1.9	166513	2 AC073787	AC073787 Mus muscu
45	170	1.9	182848	2 AC118017	AC118017 Mus muscu

## ALIGNMENTS

RESULT 1  
AX027824  
LOCUS  
DEFINITION Sequence 1 from Patent WO0034492.  
ACCESSION AX027824  
VERSION AX027824.1 GI:10188668  
KEYWORDS  
SOURCE  
ORGANISM Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 8995)  
AUTHORS Robine,S., Louvard,D., Pinto,D. and Jaisser,F.  
TITLE Regulatory sequences of the mouse villin gene - use in transgenesis  
JOURNAL Patent: WO 0034492-A 1 15-JUN-2000;

Pred. No. is the number of results predicted by chance to have a

ROBINE SYLVIE (FR) ; INST CURIE (FR) ; LOUVARD DANIEL (FR) ; PINTO DANIEL (FR) ; CENTRE NAT RECH SCIENT (FR) ; JAISSEY FREDERIC (FR)  
Location/Qualifiers  
1. 8995  
/organism="Mus sp."  
/db.xref="Enxont:10095"  
3443. 3487  
/note="exon 1"  
3489. 8981  
Intron 3489. 8981  
BASE COUNT 2275 a 2105 c 2258 g 2357 t  
ORIGIN

Query Match 100.0% Score 8995; DB 6; Length 8995;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTGTCACCAAGACACCTGCTCCAGCAGCTGGGAGGTGAGGAGAGGTGA 60  
DB 1 GATCTGTCACCAAGACACCTGCTCCAGCAGCTGGGAGGTGAGGAGAGGTGA 60  
QY 61 GAAGTTAAGTCATCTGTTTACATAGCAAGTTTCAGCCAGCTTCAGTACATGAA 120  
DB 61 GAAGTTAAGTCATCTGTTTACATAGCAAGTTTCAGCCAGCTTCAGTACATGAA 120  
QY 121 CCTTTGTTGTTGTTGTTTAAAGCATTAATAATATACCATTAAGAGGTGG 180  
DB 121 CCTTTGTTGTTGTTGTTTAAAGCATTAATAATATACCATTAAGAGGTGG 180  
QY 181 CAGTGTGCAGACACCTTTAATTCAGATTTTCAGAGGAGAGAGAGAGAGAGTCT 240  
DB 181 CAGTGTGCAGACACCTTTAATTCAGATTTTCAGAGGAGAGAGAGAGAGTCT 240  
QY 241 GAGTTGGAAGTCAAGCTTACGCAAGAGTGTCCAGAGTGGCAAGGGCTACACAGA 300  
DB 241 GAGTTGGAAGTCAAGCTTACGCAAGAGTGTCCAGAGTGGCAAGGGCTACACAGA 300  
QY 301 AACCTGTCTCAATAAACCAAGTAGTAGTAGTAATGCCATAGAGAAATTTGA 360  
DB 301 AACCTGTCTCAATAAACCAAGTAGTAGTAGTAATGCCATAGAGAAATTTGA 360  
QY 361 GTCCATTACAGATGACCATCTCTATAGATGATCTCTTGACCCAGAGTAACTATGCA 420  
DB 361 GTCCATTACAGATGACCATCTCTATAGATGATCTCTTGACCCAGAGTAACTATGCA 420  
QY 421 TGGGGAAGAGGATGGGAGTGTCTGATTAATAAAGTGTGAGGAGTGTCTGCA 480  
DB 421 TGGGGAAGAGGATGGGAGTGTCTGATTAATAAAGTGTGAGGAGTGTCTGCA 480  
QY 481 TTTGATTCATATGAAGAAGCTGATTAAGGCCCAAGAGAGTGGAACTGGAGCTGGACT 540  
DB 481 TTTGATTCATATGAAGAAGCTGATTAAGGCCCAAGAGAGTGGAACTGGAGCTGGACT 540  
QY 541 GAAGACGTGACGGCTTATTAACACTGACCTTATTAACACTGACCTTATTAACACTGACAG 600  
DB 541 GAAGACGTGACGGCTTATTAACACTGACCTTATTAACACTGACCTTATTAACACTGACAG 600  
QY 601 CGTTACGTTTGAAGTCACTTTCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 601 CGTTACGTTTGAAGTCACTTTCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 GTAGCAGAGTGGCTGCAAGAGAGTATTTAGTGAAGAGTACCTTACATATCTTT 720  
DB 661 GTAGCAGAGTGGCTGCAAGAGAGTATTTAGTGAAGAGTACCTTACATATCTTT 720  
QY 721 GCACCTATACATACACAGTGTCAAAATGCTAACTCCATAGTCCACAGATGCTGTAA 780  
DB 721 GCACCTATACATACACAGTGTCAAAATGCTAACTCCATAGTCCACAGATGCTGTAA 780  
QY 781 CTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTT 840  
DB 781 CTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTT 840  
QY 841 ATTTATTTGCTGAGAGACCATCAAGGCTTTTCACTTTGAGGACATGCTTTAC 900

DB 841 ATTTATTTGCTGAGAGACCATCAAGGCTTTTCACTTTGAGGACATGCTTTAC 900  
QY 901 TTAGGAGGCTACTTCTCCCAAGGCTTTGAACCATTTGTTTATTTATTTATTTGCT 960  
DB 901 TTAGGAGGCTACTTCTCCCAAGGCTTTGAACCATTTGTTTATTTATTTATTTGCT 960  
QY 961 GCATGAGTATGAT 1020  
DB 961 GCATGAGTATGAT 1020  
QY 1021 ATCATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
DB 1021 ATCATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
QY 1081 GAGGCAAGGCTGGGTTCTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB 1081 GAGGCAAGGCTGGGTTCTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 CTAGGCTTAAGTATCTTTTAAATATATATATATATATATATATATATATATATAT 1200  
DB 1141 CTAGGCTTAAGTATCTTTTAAATATATATATATATATATATATATATATATATAT 1200  
QY 1201 ACAGCCTTTAATCCAGACCTTGAGAGGCTGATGATGATGATGATGATGATGATGATG 1260  
DB 1201 ACAGCCTTTAATCCAGACCTTGAGAGGCTGATGATGATGATGATGATGATGATGATG 1260  
QY 1261 CTGGGCTGAGAGCTGGGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 1320  
DB 1261 CTGGGCTGAGAGCTGGGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 1320  
QY 1321 CCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
DB 1321 CCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
QY 1381 TGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
DB 1381 TGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
QY 1441 AGTGTCTTAATCTTGTGAGCCACCCCTCCACACCCCTGTTTGAAGACTTAACTTTTG 1500  
DB 1441 AGTGTCTTAATCTTGTGAGCCACCCCTCCACACCCCTGTTTGAAGACTTAACTTTTG 1500  
QY 1501 TGTAAATGTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1560  
DB 1501 TGTAAATGTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1560  
QY 1561 CTGAGCCCTGACCCACACGACTGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
DB 1561 CTGAGCCCTGACCCACACGACTGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
QY 1621 ATAGTTGATTAAGTGTGATTAATAGTCTGAGACTATGCTTGAAGGCTTGTGCTGCT 1680  
DB 1621 ATAGTTGATTAAGTGTGATTAATAGTCTGAGACTATGCTTGAAGGCTTGTGCTGCT 1680  
QY 1681 TTAGCATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
DB 1681 TTAGCATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
QY 1741 GAGGCAAGGAGAACCCACTCTGAAAGTTGTTCTGAGCTTCAATCAACTTCAAT 1800  
DB 1741 GAGGCAAGGAGAACCCACTCTGAAAGTTGTTCTGAGCTTCAATCAACTTCAAT 1800  
QY 1801 AATAGTATCAATGAT 1860  
DB 1801 AATAGTATCAATGAT 1860  
QY 1861 GAGATGCTCAGCTTCCAGAGACACTTGTGCTTTGAGAGAGACTTATGATTTCACTTCC 1920  
DB 1861 GAGATGCTCAGCTTCCAGAGACACTTGTGCTTTGAGAGAGACTTATGATTTCACTTCC 1920  
QY 1921 AGGACTCATATGCTGCTCAGAGACCATCAAGGCTTTTCACTTTGAGGACATGCTTTAC 1980



Db 1921 AGGATCATATGTGGCTCACAGCCATCTGTAATCCAGTTCCAGAGGGTTCCACACCCT 1980  
Qy 1981 CTTGTGGCTCCACAGGACCCACATACATGTACACAGACATACATGACGGCAAAACACC 2040  
Db 1981 CTTGTGGCTCCACAGGACCCACATACATGTACACAGACATACATGACGGCAAAACACC 2040  
Qy 2041 CATACACACATAAATTAATTAAGAAACCTTAAAGAGTGCATGTGTTGGTAAACATGTGCT 2100  
Db 2041 CATACACACATAAATTAATTAAGAAACCTTAAAGAGTGCATGTGTTGGTAAACATGTGCT 2100  
Qy 2101 TACACATGTGTTGAAGACATGTACAAACGACACACTGAAGAGGATCTGGGCTGGAG 2160  
Db 2101 TACACATGTGTTGAAGACATGTACAAACGACACACTGAAGAGGATCTGGGCTGGAG 2160  
Qy 2161 AGATGGCTCAGCGGTTAAGACACTGCTCTCCGAAGAGAGGTCCTGAGTTCAAT 2220  
Db 2161 AGATGGCTCAGCGGTTAAGACACTGCTCTCCGAAGAGAGGTCCTGAGTTCAAT 2220  
Qy 2221 CCTAGCAACACATGTGGCTCACACACCATCATATGAGATCTGACACCCTCTTGCT 2280  
Db 2221 CCTAGCAACACATGTGGCTCACACACCATCATATGAGATCTGACACCCTCTTGCT 2280  
Qy 2281 GCATCTGAGACAGCTCGACAGCTACAGTGTACTTGAATATCTAATTAATTAATCTTTT 2340  
Db 2281 GCATCTGAGACAGCTCGACAGCTACAGTGTACTTGAATATCTAATTAATTAATCTTTT 2340  
Qy 2341 TTTTAAAAAATGAGAGGATCTGACACCTCAAAAGAGATTGTGAGCATGTGACTCAG 2400  
Db 2341 TTTTAAAAAATGAGAGGATCTGACACCTCAAAAGAGATTGTGAGCATGTGACTCAG 2400  
Qy 2401 GGTGATTTATCTATCTGTGAGTTTCTTCCGCTTGCTGCAACTGGAGGGGAGACAG 2460  
Db 2401 GGTGATTTATCTATCTGTGAGTTTCTTCCGCTTGCTGCAACTGGAGGGGAGACAG 2460  
Qy 2461 CCCCTTTTCATTCACAAAGAGCGGTCTACATTAATTTCTGAACAAACAGCAGCTGAGT 2520  
Db 2461 CCCCTTTTCATTCACAAAGAGCGGTCTACATTAATTTCTGAACAAACAGCAGCTGAGT 2520  
Qy 2521 ATGTTTACTGCTCTGTGATATGAGACAGCGGCGGCGGCGGCGACACACACACAC 2580  
Db 2521 ATGTTTACTGCTCTGTGATATGAGACAGCGGCGGCGGCGGCGACACACACACAC 2580  
Qy 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTGGAGAGTCA 2640  
Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTGGAGAGTCA 2640  
Qy 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCCCTCTCTAAAGAGACACAGATTCC 2700  
Db 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCCCTCTCTAAAGAGACACAGATTCC 2700  
Qy 2701 AAGTGGCAGAAATCTACAGGGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2760  
Db 2701 AAGTGGCAGAAATCTACAGGGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2760  
Qy 2761 AGAGACCTACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2820  
Db 2761 AGAGACCTACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2820  
Qy 2821 CCCCTTTTCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2880  
Db 2821 CCCCTTTTCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2880  
Qy 2881 GCTCATTAAGAGTGTGCTGCTCACTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2940  
Db 2881 GCTCATTAAGAGTGTGCTGCTCACTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2940  
Qy 2941 GCCCAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000  
Db 2941 GCCCAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000  
Qy 3001 CTGTCTGTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3060  
Db 3001 CTGTCTGTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3060

Qy 3061 CAAGACCCCAAGAGTCTCTACTCCATCCCATCCAGTGCGCCCTTGCCCGGACACACC 3120  
Db 3061 CAAGACCCCAAGAGTCTCTACTCCATCCCATCCAGTGCGCCCTTGCCCGGACACACC 3120  
Qy 3121 CCCCAGACTCCCGGACACTTCTCTAGGGGCTGAGGGGTGGGACCCCTGTGGGGGTTC 3180  
Db 3121 CCCCAGACTCCCGGACACTTCTCTAGGGGCTGAGGGGTGGGACCCCTGTGGGGGTTC 3180  
Qy 3181 CTACCTCAGGTAGAGCCCAAGGTCTTAGCCGGAAGTCCACCCCATCTGAGCTGAGA 3240  
Db 3181 CTACCTCAGGTAGAGCCCAAGGTCTTAGCCGGAAGTCCACCCCATCTGAGCTGAGA 3240  
Qy 3241 GCCAAGGGGGGGGACACGGGAGCTCAGGGCTGAGGGCTGTGGGCTCTAGTTC 3300  
Db 3241 GCCAAGGGGGGGGACACGGGAGCTCAGGGCTGAGGGCTGTGGGCTCTAGTTC 3300  
Qy 3301 AGGACCTGGGACACTTCTCCACCCCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360  
Db 3301 AGGACCTGGGACACTTCTCCACCCCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3360  
Qy 3361 CTTATATGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3420  
Db 3361 CTTATATGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3420  
Qy 3421 GTCCTCAGAGCCAGCTTGGCACACACTTCTAAGATCTCCAGGTTGGGCTGCTTC 3480  
Db 3421 GTCCTCAGAGCCAGCTTGGCACACACTTCTAAGATCTCCAGGTTGGGCTGCTTC 3480  
Qy 3481 CAGACAGGTAAAGCAATTTGGGTGGGACACATGTGACACAGGTGTTGAGGGGACAG 3540  
Db 3481 CAGACAGGTAAAGCAATTTGGGTGGGACACATGTGACACAGGTGTTGAGGGGACAG 3540  
Qy 3541 GGTCTTGTCTCTCTCTGAGAGCTGTGCTTCTGTAGACACTTGGTAAGTTTGGGG 3600  
Db 3541 GGTCTTGTCTCTCTCTGAGAGCTGTGCTTCTGTAGACACTTGGTAAGTTTGGGG 3600  
Qy 3601 TGAGTAAAGGTGCTCTGAACCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3660  
Db 3601 TGAGTAAAGGTGCTCTGAACCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3660  
Qy 3661 TCAATGAAGAGGAGTTCACAGACCCCTTCTCTGAATCACTTCCCTGATCTGTAG 3720  
Db 3661 TCAATGAAGAGGAGTTCACAGACCCCTTCTCTGAATCACTTCCCTGATCTGTAG 3720  
Qy 3721 ATTCCCTGGGACCAAGGTGGCTCTGAGACTCAGATTCTAACAATCAAGACAGT 3780  
Db 3721 ATTCCCTGGGACCAAGGTGGCTCTGAGACTCAGATTCTAACAATCAAGACAGT 3780  
Qy 3781 CTTGAGACTTGGAGTCCGCTGCTATTTACTACTTCTCTCTCTCTCTCTCTCTCTCT 3840  
Db 3781 CTTGAGACTTGGAGTCCGCTGCTGATTTACTACTTCTCTCTCTCTCTCTCTCTCTCT 3840  
Qy 3841 TCATGCTTACACATCTGAATGTTCTTGTGTACCAATTCCTGACACTCTGGGA 3900  
Db 3841 TCATGCTTACACATCTGAATGTTCTTGTGTACCAATTCCTGACACTCTGGGA 3900  
Qy 3901 GGTGATCTTGGACATCTATCTGAGTGAAGCTGACGCCACGAGAGAGAGGGG 3960  
Db 3901 GGTGATCTTGGACATCTATCTGAGTGAAGCTGACGCCACGAGAGAGAGGGG 3960  
Qy 3961 AGAGTCAAGAGGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4020  
Db 3961 AGAGTCAAGAGGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4020  
Qy 4021 GGGCCCTCCATTTTTCGTTACCATGATCTAATTTATATCAGAGTGGGAGTGAAGCCA 4080  
Db 4021 GGGCCCTCCATTTTTCGTTACCATGATCTAATTTATATCAGAGTGGGAGTGAAGCCA 4080  
Qy 4081 AACCTGCCAGAAATTGGAGTCACTCAGACCAAGGTAATCTGCTCAGAAATCCCTG 4140  
Db 4081 AACCTGCCAGAAATTGGAGTCACTCAGACCAAGGTAATCTGCTCAGAAATCCCTG 4140

Oy	4141	TCAC	TTCAGGTTGGGAGAAATCTCCCTCTGGGGGCTTCCAGGCTTGGTTACGAGAGGCT	4200
Db	4141	TCAC	TTCAGGTTGGGAGAAATCTCCCTCTGGGGGCTTCCAGGCTTGGTTACGAGAGGCT	4200
Oy	4201	ATCC	TTCATAGGAGCATATGACACTAGCTATAGGGGTTACTACATTCCTCTCCAGTTAAAG	4260
Db	4201	ATCC	TTCATAGGAGCATATGACACTAGCTATAGGGGTTACTACATTCCTCTCCAGTTAAAG	4260
Oy	4261	CTGG	AACTAAAAACCACGAGCGCCAGGATTTCTCTACAGTTGTACCACAAGAACACA	4320
Db	4261	CTGG	AACTAAAAACCACGAGCGCCAGGATTTCTCTACAGTTGTACCACAAGAACACA	4320
Oy	4321	AGAC	ACATAGATATGCAAGATAGCTAGCTGGGAGAGAAGAACTTAAACCCCAAG	4380
Db	4321	AGAC	ACATAGATATGCAAGATAGCTAGCTGGGAGAGAAGAACTTAAACCCCAAG	4380
Oy	4381	GCC	CAAGTTCCGTTCCCTAGTTTCACATATGACATAGAGTGCATAGCTATGGGCTG	4440
Db	4381	GCC	CAAGTTCCGTTCCCTAGTTTCACATATGACATAGAGTGCATAGCTATGGGCTG	4440
Oy	4441	TGAC	TTCGTTGGTACATCAAGCATATGATATGCTATGCTATGTTATATCTGACAC	4500
Db	4441	TGAC	TTCGTTGGTACATCAAGCATATGATATGCTATGCTATGTTATATCTGACAC	4500
Oy	4501	TTGG	AGGCTTCGAACGAGAGATTTGCTATATGTTTGAGGCCAGCTGAGCTATAGAGCA	4560
Db	4501	TTGG	AGGCTTCGAACGAGAGATTTGCTATATGTTTGAGGCCAGCTGAGCTATAGAGCA	4560
Oy	4561	GAC	TTTGTCTTTAAGAAAAATGAAAGCCACAGATGGTGGCACAGCCCTTTAATCCA	4620
Db	4561	GAC	TTTGTCTTTAAGAAAAATGAAAGCCACAGATGGTGGCACAGCCCTTTAATCCA	4620
Oy	4621	GCAC	TTCGGAGGCGAGAGCAGGAGATTTTCTGATTTCAAGGCCACCTGGTATAGAGT	4680
Db	4621	GCAC	TTCGGAGGCGAGAGCAGGAGATTTTCTGATTTCAAGGCCACCTGGTATAGAGT	4680
Oy	4681	GAG	TTCAGGACGACCGGCTACACAGAGAAACCCCTGTTTGAAGAAACCGAGAAACAA	4740
Db	4681	GAG	TTCAGGACGACCGGCTACACAGAGAAACCCCTGTTTGAAGAAACCGAGAAACAA	4740
Oy	4741	AAC	AAAAACAAACAAACCAAAACCCCAACCCCAACCTCTCATCTCTCATCTCTC	4800
Db	4741	AAC	AAAAACAAACAAACCAAAACCCCAACCCCAACCTCTCATCTCTCTCATCTCTC	4800
Oy	4801	TAG	CTCTGTCTGTCTAGTGTGTAGATTTGGGAGCTTCACTATATATTAATAGGCC	4860
Db	4801	TAG	CTCTGTCTGTCTAGTGTGTAGATTTGGGAGCTTCACTATATATTAATAGGCC	4860
Oy	4861	TTTT	TATTCACCTGTCTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGACCCCTGAGA	4920
Db	4861	TTTT	TATTCACCTGTCTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGACCCCTGAGA	4920
Oy	4921	AAG	TATTCACCTGTCTAGAGACGAGAAAGTTTCTGGGAAGCTTCTGGAGAGTGTCTCCGAT	4980
Db	4921	AAG	TATTCACCTGTCTAGAGACGAGAAAGTTTCTGGGAAGCTTCTGGAGAGTGTCTCCGAT	4980
Oy	4981	CAG	ACTACTGTTCAGAGCAGAGAAGAGGTTGGAATATGTTGGTGCACAGCACTT	5040
Db	4981	CAG	ACTACTGTTCAGAGCAGAGAAGAGGTTGGAATATGTTGGTGCACAGCACTT	5040
Oy	5041	GGA	CAGAGAGAGAGAGGAGGAGGATTCACAGATTTCTGACATGTAGCTGACTTTTGGT	5100
Db	5041	GGA	CAGAGAGAGAGAGGAGGAGGATTCACAGATTTCTGACATGTAGCTGACTTTTGGT	5100
Oy	5101	TCT	CTGGGTGACAAAGTCTCCCAAGGATATAGGCTTTACAAAAGGGACACAGGGGTGAGCC	5160
Db	5101	TCT	CTGGGTGACAAAGTCTCCCAAGGATATAGGCTTTACAAAAGGGACACAGGGGTGAGCC	5160
Oy	5161	AAT	GATTCAGGTTGAGGGACATATCAGGCCAGGGTCTCTGCTGGCAAGCTAAAGATG	5220
Db	5161	AAT	GATTCAGGTTGAGGGACATATCAGGCCAGGGTCTCTGCTGGCAAGCTAAAGATG	5220
Oy	5221	AGAG	CCCTTAAACCTCTCTGAGATTTAGGGAGACAGAGAGAGCTGAGAGATCTCTCTTA	5280

Db	5221	AGAGCCCTTAACCCCTCCCTGAAAGTTTAGGGGAGACAGAGAGCTGAGGAGATCCCTTGA		5280
OY	5281	GGGTGAAGAGAGAGTATCTCTCTCGACCAACATAGCTAGAGACAGAGACAGTTGGACAG		5340
Db	5281	GGGTGAAGAGAGAGTATCTCTCTGACCAACATAGCTAGAGACAGAGAGATTTGGACAG		5340
OY	5341	TTTACCCTCAGAACACAGCCATCCCTCTTGGCTTAAGAGAGCTGGGCCCTTTCTGT		5400
Db	5341	TTTACCCTCAGAACACAGCCATCCCTCTTGGCTTAAGAGAGCTGGGCCCTTTCTGT		5400
OY	5401	AAGAACTCTTACTTTTCTTCAGAGAGAGGACAGAGCTTTGTCCTCTCTTGGTCA		5460
Db	5401	AAGAACTCTTACTTTTCTTCAGAGAGAGGACAGAGCTTTGTCCTCTCTTGGTCA		5460
OY	5461	TAAACACCCCGTGTGTAAATTAAGTTTATTTACTGTGAGTTTGCTCCAGACAGTCCA		5520
Db	5461	TAAACACCCCGTGTGTAAATTAAGTTTATTTACTGTGAGTTTGCTCCAGACAGTCCA		5520
OY	5521	TCTGGTAGACCTCTGCTCACTACCAAGATATGGCCACATTCCTTACCAGAGAG		5580
Db	5521	TCTGGTAGACCTCTGCTCACTACCAAGATATGGCCACATTCCTTACCAGAGAG		5580
OY	5581	TGCAGAGAGAGCCTTGTAGAGAAAGGTTACAGTACAAACATGGCCAGATTAACAA		5640
Db	5581	TGCAGAGAGAGCCTTGTAGAGAAAGGTTACAGTACAAAGATGGCCAGATTAACAA		5640
OY	5641	ACTACTATCCCTTTTACCAGAAATTTGGTTTGGTGAACACAGAGAGGGGTGTGACTGTAT		5700
Db	5641	ACTACTATCCCTTTTACCAGAAATTTGGTTTGGTGAACACAGAGAGGGGTGTGACTGTAT		5700
OY	5701	GTTGT		5760
Db	5701	GTTGT		5760
OY	5761	CTTGGGGACTTTTCATGCTAAAGATATCTGATATTGGCCCATGCCACAGGGGTAT		5820
Db	5761	CTTGGGGACTTTTCATGCTAAAGATATCTGATATTGGCCCATGCCACAGGGGTAT		5820
OY	5821	TGGGGAGAGTCAGGCTTCTGCAAAACAGTAACTGCCAAGATGGATGGTGGCTGAA		5880
Db	5821	TGGGGAGAGTCAGGCTTCTGCAAAACAGTAACTGCCAAGATGGATGGTGGCTGAA		5880
OY	5881	TCACCAAGGGGACAGCTGATCAGAGTGTGACAGAACTCAACAGATTAAGCACCCCTGTGG		5940
Db	5881	TCACCAAGGGGACAGCTGATCAGAGTGTGACAGAACTCAACAGATTAAGCACCCCTGTGG		5940
OY	5941	GCTCAGAGAGGAGTTTACAAGAGGTAAAGGCCAAGCCATTATTATTCGAAGACATGAC		6000
Db	5941	GCTCAGAGAGGAGTTTACAAGAGGTAAAGGCCAAGCCATTATTATTCGAAGACATGAC		6000
OY	6001	TCAAAATCAAAAGTCCAGAGAGAGATTAGCTGTGAGAGATGGGGCTGCACGTGGGACAC		6060
Db	6001	TCAAAATCAAAAGTCCAGAGAGAGATTAGCTGTGAGAGATGGGGCTGCACGTGGGACAC		6060
OY	6061	TGACCTTGCACCTTATTAAGTCACTAAGGCCAAGAGACGTACAGAGAGGTACTGGGTCTTA		6120
Db	6061	TGACCTTGCACCTTATTAAGTCACTAAGGCCAAGAGACGTACAGAGAGGTACTGGGTCTTA		6120
OY	6121	CTCAGCTTGGAGAGGACGTGTGAAGATGGGTACTCTCATCTGTATGAGAGGGCTGAG		6180
Db	6121	CTCAGCTTGGAGAGGACGTGTGAAGATGGGTACTCTCATCTGTATGAGAGGGCTGAG		6180
OY	6181	CACCAACAGGTACAAAGTTCCTCGTGTCTCATATGCCAGGATTCCTGGCCAGTTTCAAG		6240
Db	6181	CACCAACAGGTACAAAGTTCCTCGTGTCTCATATGCCAGGATTCCTGGCCAGTTTCAAG		6240
OY	6241	GACTAAGAGACTCATCTGTGTGAAGCAAAAGTATCCAAAGCCCTAAGCCCCATTTTGGTCT		6300
Db	6241	GACTAAGAGACTCATCTGTGTGAAGCAAAAGTATCCAAAGCCCTAAGCCCCATTTTGGTCT		6300
OY	6301	AATTAATCAGAACCCCTGGGGATGACAGGCTGTGACAGCAGAGAGCTTTTAAAGCTC		6360

Db 6301 AATTAAATCAGAACCCCTGGGGATGCAAGCTCTGAGCAGCAGAGACTTTTAAAAAGCTC 6360  
Qy 6361 CCAGGTGATTCGATACACAGCTGGAAACAACAGACTACAGGTTCACAACGAAGAAGGC 6420  
Db 6361 CCAAGTATCTGATACACAGCTGGAAACAACAGCTACAGGTTCACAACGAAGAAGGC 6420  
Qy 6421 AAAGCTAGGGAAGACTTTGGGATGGGAGACCTTCTCCAGGCGAGTAGATGAGGCTGGTT 6480  
Db 6421 AAAGCTAGGGAAGACTTTGGGATGGGAGACCTTCTCCAGGCGAGTAGATGAGGCTGGTT 6480  
Qy 6481 AGCAGGTGGCAGCCTCTCTGCTGCTGTCAATATAGCTATATCCATCCATCCATCCAT 6540  
Db 6481 AGCAGGTGGCAGCCTCTCTGCTGCTGTCAATATAGCTATATCCATCCATCCATCCAT 6540  
Qy 6541 ACACCCACCCATCCATTTATATGACCCCATCCCTTCATCCATCCATCCATCCATCCAC 6600  
Db 6541 ACACCCACCCATCCATTTATGACCCCATCCCTTCATCCATCCATCCATCCATCCAC 6600  
Qy 6601 CCAGGCATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Db 6601 CCAGGCATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Qy 6661 CATTATTCACAGAGAACTGGTATTTGACTAAATGTGGAGATTAAATTAATTTTGA 6720  
Db 6661 CATTATTCACAGAGAACTGGTATTTGACTAAATGTGGAGATTAAATTAATTTTGA 6720  
Qy 6721 AGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
Db 6721 AGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
Qy 6781 TGGCAATCGAGAAAGGTTTGGGATGGTGTCTTCTCCACGCGTGGGTTCTGGGGA 6840  
Db 6781 TGGCAATCGAGAAAGGTTTGGGATGGTGTCTTCTCCACGCGTGGGTTCTGGGGA 6840  
Qy 6841 TTGAACCTCAAAATATCGGGCTGGTGGCAAGTGTCTTACACGAGCCATTTTGTGACA 6900  
Db 6841 TTGAACCTCAAAATATCGGGCTGGTGGCAAGTGTCTTACACGAGCCATTTTGTGACA 6900  
Qy 6901 CATCATTTATTTAGAAAGCATCTTATGTATGATCCAGGCTGGCCTTCAGCTTGTATGTCG 6960  
Db 6901 CATCATTTATTTAGAAAGCATCTTATGTATGATCCAGGCTGGCCTTCAGCTTGTATGTCG 6960  
Qy 6961 CCAGGATGACCTTTAACTCCTGCTCTCCAGCCCTCCAGGAGTGGATTTTACAGGT 7020  
Db 6961 CCAGGATGACCTTTAACTCCTGCTCTCCAGCCCTCCAGGAGTGGATTTTACAGGT 7020  
Qy 7021 GTTCAACTGTGTAATGCCCTTTAATCCAGCACTCTGTGGGGGGGGGGGAGGCGGATC 7080  
Db 7021 GTTCAACTGTGTAATGCCCTTTAATCCAGCACTCTGTGGGGGGGGGGGAGGCGGATC 7080  
Qy 7081 CCTGAGTTGGAGGCGAGTTTGGTGTACAGAGTTTCAGATACCTTGGGGCTATACAGGAA 7140  
Db 7081 CCTGAGTTGGAGGCGAGTTTGGTGTACAGAGTTTCAGATACCTTGGGGCTATACAGGAA 7140  
Qy 7141 ACCCTATCCCAACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7200  
Db 7141 ACCCTATCCCAACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 7200  
Qy 7201 TTAGAGATATTTAGAGGTAGTAGGGCTGGTGGAGGAGAGTCAATGCTTTCTTTGTATTT 7260  
Db 7201 TTAGAGATATTTAGAGGTAGTAGGGCTGGTGGAGGAGAGTCAATGCTTTCTTTGTATTT 7260  
Qy 7261 ATATATAGTAAAGTACTCACAAGATGATATTTCTATCTATCTATCTATCTATCTATCTATC 7320  
Db 7261 ATATATAGTAAAGTACTCACAAGATGATATTTCTATCTATCTATCTATCTATCTATCTATC 7320  
Qy 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380  
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380  
Qy 7381 CTGCTTTGACTCTGAATGCTCTATTTCTGGGTCAACTCTTACCCCTAGTGTGGGTTT 7440  
Db 7381 CTGCTTTGACTCTGAATGCTCTATTTCTGGGTCAACTCTTACCCCTAGTGTGGGTTT 7440

Qy 7441 ACCAAGACCCAGACATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 7500  
Db 7441 ACCAAGACCCAGACATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 7500  
Qy 7501 GGACTCAGGGCTCTTGTGATCTTAAGCAAGCTCTGTGCACAGAGCTGCACCTCCAGTCC 7560  
Db 7501 GGACTCAGGGCTCTTGTGATCTTAAGCAAGCTCTGTGCACAGAGCTGCACCTCCAGTCC 7560  
Qy 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTGATATTTGCAAGCGCTATGTAGCTCTCTC 7620  
Db 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTGATATTTGCAAGCGCTATGTAGCTCTCTC 7620  
Qy 7621 CACCTCCAGTTCCAGACACTTCTGTGTCATCCAGTGGGGGGGGAACCTGTGTCACACA 7680  
Db 7621 CACCTCCAGTTCCAGACACTTCTGTGTCATCCAGTGGGGGGGGAACCTGTGTCACACA 7680  
Qy 7681 GTGCCCTGTCCCTGTCTGTGACCTACATATTTGCTGTGAACAGTTTATGTAAATG 7740  
Db 7681 GTGCCCTGTCCCTGTCTGTGACCTACATATTTGCTGTGAACAGTTTATGTAAATG 7740  
Qy 7741 GGATGCGTTCCTGTGTATTTCTTTTATGGCTGGCCCTTTATCTTACAGCAGTTTGCTTG 7800  
Db 7741 GGATGCGTTCCTGTGTATTTCTTTATGGCTGGCCCTTTATCTTACAGCAGTTTGCTTG 7800  
Qy 7801 GGCCATGTGTACGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
Db 7801 GGCCATGTGTACGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
Qy 7861 TGTGTGATTAACACATTTCTGTTCATTTACTGATGGAATTTGTGGCCCCCACCAC 7920  
Db 7861 TGTGTGATTAACACATTTCTGTTCATTTACTGATGGAATTTGTGGCCCCCACCAC 7920  
Qy 7921 CCTTTTTTTTTTATTTGAGACAAGCTTCTGTATCTGTATCTTGCATCTTGGCTGTCTT 7980  
Db 7921 CCTTTTTTTTTTATTTGAGACAAGCTTCTGTATCTGTATCTTGCATCTTGGCTGTCTT 7980  
Qy 7981 GAGCTCACTGTGTAGACAGGCTGTGAGGCTGTCTTCCACTTTTGAACCTCTGTGAC 8040  
Db 7981 GAGCTCACTGTGTAGACAGGCTGTGAGGCTGTCTTCCACTTTTGAACCTCTGTGTGAC 8040  
Qy 8041 AGAGTAGCATGATGATCAAAAGCAATTTCTGTTTGGTTGGTTTGTATTTTACATTTGTGTG 8100  
Db 8041 AGAGTAGCATGATGATCAAAAGCAATTTCTGTTTGGTTGGTTTGTATTTTACATTTGTGTG 8100  
Qy 8101 TGTATGCGTGTATATGTGATGTTTGTGTCTTCAAGTGTCTCAATGTGTGACCTGTGTG 8160  
Db 8101 TGTATGCGTGTATATGTGATGTTTGTGTCTTCAAGTGTCTCAATGTGTGACCTGTGTG 8160  
Qy 8161 TGGGACAGAGAAACAACCGAGTGCATCTTCCAGATCTACAGCATCTTGTATATATGTA 8220  
Db 8161 TGGGACAGAGAAACAACCGAGTGCATCTTCCAGATCTACAGCATCTTGTATATATGTA 8220  
Qy 8221 TGTATATATGTATGTATATTTATGTGTGCGCAAGTATGCAAGTATTTTGTGGAGTTTTCAC 8280  
Db 8221 TGTATATATGTATGTATATTTATGTGTGCGCAAGTATGCAAGTATTTTGTGGAGTTTTCAC 8280  
Qy 8281 CTTCCTTTGTGGGCTCTCCGCAATTAACCTCAGCTCTCGGCTAGTGAAGTATCCCTTCA 8340  
Db 8281 CTTCCTTTGTGGGCTCTCCGCAATTAACCTCAGCTCTCGGCTAGTGAAGTATCCCTTCA 8340  
Qy 8341 CTGATGAGCCATCTGCTGTGCGCTGTGCGCACTCTCTCTTATTTTCCCAATGGGACTA 8400  
Db 8341 CTGATGAGCCATCTGCTGTGCGCTGTGCGCACTCTCTCTTATTTTCCCAATGGGACTA 8400  
Qy 8401 CGCAGTCACTGGGCTTAAAGCTTCAACAAGTCACTGAGAGTGGTAAAGCAGGAGACTTCAG 8460  
Db 8401 CGCAGTCACTGGGCTTAAAGCTTCAACAAGTCACTGAGAGTGGTAAAGCAGGAGACTTCAG 8460  
Qy 8461 GGATATGCTGGGCTCTGCTGCTCAGAGTGTGAATTTACAGGCATATCATCTGCTGGAAG 8520  
Db 8461 GGATATGCTGGGCTCTGCTGCTCAGAGTGTGAATTTACAGGCATATCATCTGCTGGAAG 8520

Qy 8521 ATTTTAACTGAATCTGAGATAGAGAGGACCTCTACCAATGAGGGCTCTTTTGT 8580  
 |||||||  
 Db 8521 ATTTTAACTGAATCTGAGATAGAGAGGACCTCTACCAATGAGGGCTCTTTTGT 8580  
 Qy 8581 GTTTTGGTTTGGCTTCTCTGCAATAGATAGAGAGCTGTAAGTGTAGCCCTGGCTAC 8640  
 |||||||  
 Db 8581 GTTTTGGTTTGGCTTCTCTGCAATAGATAGAGAGCTGTAAGTGTAGCCCTGGCTAC 8640  
 Qy 8641 ATTAACATCTTCTCTCAAAAAGCCTATAGAGTGGAGGCTGCGAGCTTAAAGAGACCTT 8700  
 |||||||  
 Db 8641 ATTAACATCTTCTCTCAAAAAGCCTATAGAGTGGAGGCTGCGAGCTTAAAGAGACCTT 8700  
 Qy 8701 AAGCCGCGTGTGATAGACACAGAGATAGCCTGCATATATAGACAGACCTTGTTCAAAA 8760  
 |||||||  
 Db 8701 AAGCCGCGTGTGATAGACACAGAGATAGCCTGCATATATAGACAGACCTTGTTCAAAA 8760  
 Qy 8761 ACATGAGAGGAGGGGTATGTTTTAACTGCTGGCTGTGTAAACAGGACCTAAGGACCAA 8820  
 |||||||  
 Db 8761 ACATGAGAGGAGGGGTATGTTTTAACTGCTGGCTGTGTAAACAGGACCTAAGGACCAA 8820  
 Qy 8821 TGTAGACATTTTGTAGTAAAGAGATCATCATCAAAAGCGGGTGGAGGTTAGGTTGG 8880  
 |||||||  
 Db 8821 TGTAGACATTTTGTAGTAAAGAGATCATCATCAAAAGCGGGTGGAGGTTAGGTTGG 8880  
 Qy 8881 ACTACAGTGGTCAAGACCCCATAGGAGCCAGTTTCCCTTCTTCTGCGCCCTCAAGC 8940  
 |||||||  
 Db 8881 ACTACAGTGGTCAAGACCCCATAGGAGCCAGTTTCCCTTCTTCTGCGCCCTCAAGC 8940  
 Qy 8941 CTGGCTCGACGGGCACTGCTCTACATGCTTCTCTCTAGGCTGCTCCACCATG 8995  
 |||||||  
 Db 8941 CTGGCTCGACGGGCACTGCTCTACATGCTTCTCTCTAGGCTGCTCCACCATG 8995

## RESULT 2

AC098570

LOCUS AC098570 192060 bp DNA linear HTG 20-AUG-2002

DEFINITION Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered

pieces.

AC098570.2 GI:22325297

AC098570.2 GI:22325297  
 HTG: HTGS\_PHASE1: HTGS\_DRAFT: HTGS\_FULFILLTOP.  
 house mouse.

SOURCE Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 192060)

Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-278N11  
 Unpublished

2 (bases 1 to 192060)

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhvalter, B.,  
 Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,  
 Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
 Girdya, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hago, B., Heaford, A., Horton, L., Hulme, W., Illiev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kellis, C., Larocque, K.,  
 Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Meldrum, J.,  
 Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norou, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Plerie, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Topham, K., Travers, M., Trivis, N., Trigglio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission

## JOURNAL

REFERENCE

AUTHORS

## COMMENT

JOURNAL

COMMENT

Submitted (24-OCT-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 192060)  
 Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhvalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choe, Y., Colangelo, M.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
 Gaidyna, S., Girdya, S., Graham, L., Grand-Pierre, N., Hago, B.,  
 Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C., Kamat, A.,  
 Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,  
 McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mienga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norou, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Plerie, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
 Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 20, 2002 this sequence version replaced g1:16356898.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www.seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L14796  
 Center clone name: 278\_N\_11  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 190464 bases at least Q40  
 Consensus quality: 191061 bases at least Q30  
 Consensus quality: 191345 bases at least Q20  
 Insert size: 163000; agarose-fp  
 Insert size: 191560; sum-of-contigs  
 Quality coverage: 12.4 in Q20 bases; agarose-fp  
 Quality coverage: 10.5 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 4921: contig of 4921 bp in length  
 \* 4922 5021: gap of 100 bp  
 \* 5022 15200: contig of 10179 bp in length  
 \* 15201 15300: gap of 100 bp  
 \* 15301 78692: contig of 63392 bp in length  
 \* 78693 78792: gap of 100 bp  
 \* 78793 116118: contig of 37326 bp in length  
 \* 116119 116218: gap of 100 bp  
 \* 116219 174491: contig of 58273 bp in length  
 \* 174492 174591: gap of 100 bp  
 \* 174592 192060: contig of 17469 bp in length.  
 Location/Qualifiers  
 1. 192060  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-278N11"  
 /clone\_1ib="RP23 Female Mouse BAC"

```
misc_feature 1..4921
              /note="assembly_fragment
              clone_end:SP6
              vector_side:left"
misc_feature 5022..15200
              /note="assembly_fragment"
misc_feature 15301..78692
              /note="assembly_fragment"
misc_feature 78793..116118
              /note="assembly_fragment"
misc_feature 116219..174491
              /note="assembly_fragment"
misc_feature 174592..192060
              /note="assembly_fragment
              clone_end:T7
              vector_side:right"
BASE COUNT 49668 a 46807 c 46516 g 48562 t 507 others
ORIGIN
Query Match 97.4% Score 8762.2; DB 2; Length 192060;
Best Local Similarity 98.2% Pred. No. 0;
Matches 8982; Conservative 0; Mismatches 3; Indels 158; Gaps 6;
QY 1 GATCTGTGACACGACGACGCTGTGCTCCAGCACTGGGAGGTGGAGGAGAGGTCA 60
    |||||||
Db 40161 GATCTGTGACACGACGACGCTGTGCTCCAGCACTGGGAGGTGGAGGAGAGGTCA 40220
QY 61 GAAGTTTAAGGTCACTCTTGTGTTACATAGCAAGGTTTCAGCCAGCTTCAGTACATGAAA 120
    |||||||
Db 40221 GAAGTTTAAGGTCACTCTTGTGTTACATAGCAAGGTTTCAGCCAGCTTCAGTACATGAAA 40280
QY 121 CCTTGTGTTGTTGTTGTTGTTGTTTAAAGCATTAAATATATCCATAAGAGAGTGG 180
    |||||||
Db 40281 CCTTGTGTTGTTGTTGTTGTTGTTTAAAGCATTAAATATATCCATAAGAGAGTGG 40340
QY 181 CAGTGTGGACAGACACCTTTAATTCATATTCAGAGAGAGCAAGACAGACAGATCTCTGT 240
    |||||||
Db 40341 CAGTGTGGACAGACACCTTTAATTCATATTCAGAGAGAGCAAGACAGACAGATCTCTGT 40400
QY 241 GAGTTTGAAGTCAGCTAGTCTGCAAGAGCTAGTTCCAGAGTGGCAAGGGCTTACACAGAGA 300
    |||||||
Db 40401 GAGTTTGAAGTCAGCTAGTCTGCAAGAGCTAGTTCCAGAGTGGCAAGGGCTTACACAGAGA 40460
QY 301 AACCTGTCTCATTAACCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 360
    |||||||
Db 40461 AACCTGTCTCATTAACCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 40520
QY 361 GTCCATTAGAGTGAACCATCTTAATAGATGATTTCTTGAACCCAGTAAGCTAATGTCA 420
    |||||||
Db 40521 GTCCATTAGAGTGAACCATCTTAATAGATGATTTCTTGAACCCAGTAAGCTAATGTCA 40580
QY 421 TGGGGAAGAGGGAGTGAAGCTGTCTAGATTTAAAGAGTCTGAGGGAGTGCCTATTTCTCAA 480
    |||||||
Db 40581 TGGGGAAGAGGGAGTGAAGCTGTCTAGATTTAAAGAGTCTGAGGGAGTGCCTATTTCTCAA 40640
QY 481 TTTGATTCATATGAAGAAGGCTGATAAGGCCCAAGAAGTGGAACTGGGAGTCTGGAGCT 540
    |||||||
Db 40641 TTTGATTCATATGAAGAAGGCTGATAAGGCCCAAGAAGTGGAACTGGGAGTCTGGAGCT 40700
QY 541 GAAGACGTGACGGCTTTTAAACACTGGCACTTATTAACACTTATTAACACTGGCACAGG 600
    |||||||
Db 40701 GAAGACGTGACGGCTTTTAAACACTGGCACTTATTAACACTTATTAACACTGGCACAGG 40760
QY 601 CGTTTCAGGTTTGAAGATCACTTTCAAACACAGAAACAGAAAGTGTGCTGCTCTCAGC 660
    |||||||
Db 40761 CGTTTCAGGTTTGAAGATCACTTTCAAACACAGAAACAGAAAGTGTGCTGCTCTCAGC 40820
QY 661 GTAGGACACACTGGCTGTGAGAAAGATGATTTTATGAAAGTACCTTCACATATCTTT 720
    |||||||
Db 40821 GTAGGACACACTGGCTGTGAGAAAGATGATTTTATGAAAGTACCTTCACATATCTTT 40880
QY 721 GCACCTTATACATACACGCTGTCAATGTGTAACTCCCTAGTCCACAGATGGCTTTACA 780
    |||||||
```

```
Db 40881 GCACCTTATACATACACGCTGTCAATGTGTAACTCCCTAGTCCACAGATGGCTTTACA 40940
QY 781 CTCGTTTCTGCTTTCATCTGGTTGACATTTTGTGAGAACCAAGAAATAGAAATGGGT 840
    |||||||
Db 40941 CTCGTTTCTGCTTTCATCTGGTTGACATTTTGTGAGAACCAAGAAATAGAAATGGGT 41000
QY 841 ATTTATTTTCTGCTGAGAGACACCATCCAGGGCTTTTACATTTTACAGCAGCATGGTTTAC 900
    |||||||
Db 41001 ATTTATTTTGTGCTGAGAGACACCATCCAGGGCTTTTACATTTTACAGCAGCATGGTTTAC 41060
QY 901 TTAAGGCGCTACTTCTCCAAACGTTTGAACCATTTGTTTATTTTACTTATTTTGTGT 960
    |||||||
Db 41061 TTAAGGCGCTACTTCTCCAAACGTTTGAACCATTTGTTTATTTTACTTATTTTGTGT 41120
QY 961 GCATGAGTAGGCAATGATATACGTATATAGAGATCAGATGAGGCTTACCTCAAA 1020
    |||||||
Db 41121 GCATGAGTAGGCAATGATATACGTATATAGAGATCAGATGAGGCTTACCTCAAA 41180
QY 1021 ATCATTTGCAATCCCGACAGATGAAGTCAACCGACGTTGTATATGTGGAGCTG 1080
    |||||||
Db 41181 ATCATTTGCAATCCCGACAGATGAAGTCAACCGACGTTGTATATGTGGAGCTG 41240
QY 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGCAGCAGTGGCTTAAACATGGAGCAGCTCT 1140
    |||||||
Db 41241 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGCAGCAGTGGCTTAAACATGGAGCAGCTCT 41300
QY 1141 CTAGGCTTAAGGTAAATCTTTTAAATATATATATATATATATATATATATATATATATATAT 1200
    |||||||
Db 41301 CTAGGCTTAAGGTAAATCTTTTAAATATATATATATATATATATATATATATATATATATAT 41360
QY 1201 ACAGGCTTTTATCCGACAGCTTGAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1260
    |||||||
Db 41361 ACAGGCTTTTATCCGACAGCTTGAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 41420
QY 1261 CTGGGGTGCAGAGCTTGGCCCTGTTTTTTTCTTTTCTTTATGTCAGTGGTCTTCA 1320
    |||||||
Db 41421 CTGGGGTGCAGAGCTTGGCCCTGTTTTTTTCTTTTCTTTATGTCAGTGGTCTTCA 41480
QY 1321 CCTGCGTGTATGTCGTCGACAGAGGCTGACATCCCTTGGAGCTGGAATTAAGACAGTTG 1380
    |||||||
Db 41481 CCTGCGTGTATGTCGTCGACAGAGGCTGACATCCCTTGGAGCTGGAATTAAGACAGTTG 41540
QY 1381 TGTATACGCTGCGCTTACAGATGCTGGAATTAAGACCCAGGCTGCTCCCTGAGAACAGGCG 1440
    |||||||
Db 41541 TGTATACGCTGCGCTTACAGATGCTGGAATTAAGACCCAGGCTGCTCCCTGAGAACAGGCG 41600
QY 1441 AGTGCCTTTAATCTTGAAGCCACCCCTCCAAACCTGCTTTTGAAGACTTCTTAACCTTTTG 1500
    |||||||
Db 41601 AGTGCCTTTAATCTTGAAGCCACCCCTCCAAACCTGCTTTTGAAGACTTCTTAACCTTTTG 41660
QY 1501 TGTATATGTGGAGACTGAGTGTGCTTGAACCTTGAACAGTGTGTGCTGCGCTTACATCA 1560
    |||||||
Db 41661 TGTATATGTGGAGACTGAGTGTGCTTGAACCTTGAACAGTGTGTGCTGCGCTTACATCA 41720
QY 1561 CTGAGCCCGGTACCCACAGACATGAGTATACGTTTAAAGGGGAACACTTAACATATGACA 1620
    |||||||
Db 41721 CTGAGCCCGGTACCCACAGACATGAGTATACGTTTAAAGGGGAACACTTAACATATGACA 41780
QY 1621 ATAGTTGATAGAGTTTGAATATATAGTCTGACCTTATGCTTATGAGCTTATGCTGTTC 1680
    |||||||
Db 41781 ATAGTTGATAGAGTTTGAATATATAGTCTGACCTTATGCTTATGAGCTTATGCTGTTC 41840
QY 1681 TTAGCATGTGCTGTGAGAGATAGAAAAATGAAGACTTGAAGTCTAGTCTCTGGAACCCACA 1740
    |||||||
Db 41841 TTAGCATGTGCTGTGAGAGATAGAAAAATGAAGACTTGAAGTCTAGTCTCTGGAACCCACA 41900
QY 1741 GAGGAGGCGAAGACCCCTCTGAAAGTGTGCTGAGCTTCAATACACTTCAACTCAAT 1800
    |||||||
Db 41901 GAGGAGGCGAAGACCCCTCTGAAAGTGTGCTGAGCTTCAATACACTTCAACTCAAT 41960
QY 1801 AATAGTTTACATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
    |||||||
Db 41961 AATAGTTTACATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 42020
```

OY	1861	GAGATGGCTCAGCTTCCAGAGAGACCTTGCTGCTCTTGCGAGAGAGACCTAGATTCAAGTCCC	1920
Db	42021	GAGATGGCTCAGCTTCCAGAGAGACCTTGCTGCTCTTGCGAGAGAGACCTAGATTCAAGTCCC	42080
OY	1921	AGGACTCATATGGTGGGCTCACAGCCCATCGTAAATCCAGTTTCCAGAGGTTCCACACCTT	1980
Db	42081	AGGACTCATATGGTGGGCTCACAGCCCATCGTAAATCCAGTTTCCAGAGGTTCCACACCTT	42140
OY	1981	CTTCTGGCCCTCCACAGGAGCCACATACATAGTACACAGACATACATATGCGAGCAAAACAC	2040
Db	42141	CTTCTGGCCCTCCACAGGAGCCACATACATAGTACACAGACATACATATGCGAGCAAAACAC	42200
OY	2041	CATACACACATTAATTAATTAAGGAACTTTAAAGGTGCATGTTGGTAAACATTTGTGCT	2100
Db	42201	CATACACACATTAATTAATTAAGGAACTTTAAAGGTGCATGTTGGTAAACATTTGTGCT	42260
OY	2101	TACACATGCTGATTCAAGACATGTACAACGCCACACACTGAGAGAGGATCTGGGGCTGGAG	2160
Db	42261	TACACATGCTGATTCAAGACATGTACAACGCCACACACTGAGAGAGGATCTGGGGCTGGAG	42320
OY	2161	AGATGGCTCAGGGGTTAAGAGACACAGCTGCCTTCCGAGGAAGGTCCGTGATTCAAAT	2220
Db	42321	AGATGGCTCAGGGGTTAAGAGACACAGCTGCCTTCCGAGGAAGGTCCGTGATTCAAAT	42380
OY	2221	CCTACCAACCAACATGGTGGCTCACCAACCATCCATATGATGATCTACACCCCTCTTGCT	2280
Db	42381	CCTACCAACCAACATGGTGGCTCACCAACCATCCATATGATGATCTACACCCCTCTTGCT	42440
OY	2281	GCATCTGAAGACGCTGCAGAGCTACAGTGTACTTATATCTAATTAATTAATCTTTT	2340
Db	42441	GCATCTGAAGACGCTGCAGAGCTACAGTGTACTTATATCTAATTAATTAATCTTTT	42500
OY	2341	TTTAAAAAATGAAGGAGTCTGAGACACCCCAAAAGAGATTATGACAGAGTCACTCACG	2400
Db	42501	TTTAAAAAATGAAGGAGTCTGAGACACCCCAAAAGAGATTATGACAGAGTCACTCACG	42560
OY	2401	GGTGAATTATCTTCCCTGGAGTTTTTCTTCTCCGCTTGGCTTGCAACTGGGTTGACAGACG	2460
Db	42561	GGTGAATTATCTTCCCTGGAGTTTTTCTTCTCCGCTTGGCTTGCAACTGGGTTGACAGACG	42620
OY	2461	CCCCCTTTTCATTACAGAAGCGGGTGTACATTTATTTCTGAACAAACAGCACCTGCAGT	2520
Db	42621	CCCCCTTTTCATTACAGAAGCGGGTGTACATTTATTTCTGAACAAACAGCACCTGCAGT	42680
OY	2521	ATGTTTACTGTCCCTGTGCATATGACAGCGGCGCGCGCGCG - -CACACACACACAC	2578
Db	42681	ATGTTTACTGTCCCTGTGCATATGACAGCGGCGCGCGCGCGCGCACACACACACAC	42740
OY	2579	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAGGT	2638
Db	42741	ACACACACACACACACACACACACACACATTCAGTCTCCAGTCTCCAGAGCTCTTGGAAGGT	42800
OY	2639	CAAGAAGAGGCTGCCCTCAACACAGATCTTCATCTTTCCCTCTAAAGGAGACACGATT	2698
Db	42801	CAAGAAGAGGCTGCCCTCAACACAGATCTTCATCTTTCCCTCTAAAGGAGACACGATT	42860
OY	2699	CCAAGGTGGCANAABATCTACAGGGGGGCGCAGGCGAGGAGGGGGGANAAGCAGGCATAGTTT	2758
Db	42861	CCAAGGTGGCANAABATCTACAGGGGGGCGCAGGCGAGGAGGGGGGANAAGCAGGCATAGTTT	42920
OY	2759	CCAGAGACCTTACAGAGAGGGGCGCAGAAGCAGATCCCAAGTCCAGGGCGAGGAGGTGGA	2818
Db	42921	CCAGAGACCTTACAGAGAGGGGCGCAGAAGCAGATCCCAAGTCCAGGGCGAGGAGGTGGA	42980
OY	2819	GGCCCTTGTTCGAGAGAGAGGCGAGGCGCAGAACAGGGTTTCAAGGCAACAGGTTTATGG	2878
Db	42981	GGCCCTTGTTCGAGAGAGAGGCGAGGCGCAGAACAGGGTTTCAAGGCAACAGGTTTATGG	43040
OY	2879	CAGCTCATTAAGTGGAGGTGCTGGCTCACTAGAAAGAGGAGAAAGAGGAAAGGCCCTT	2938
Db	43041	CAGCTCATTAAGTGGAGGTGCTGGCTCACTAGAAAGAGGAGAAAGAGGAAAGGCCCTT	43100

QY	2939	GTGCCACTGACCGAGGTCATGCTGAGTAGAGAGATCTGCAAGGGGTGCCAGAGACCCC	2998
Db	43101	GTGCCACTGACCGAGGTCATGCTGAGTAGAGAGATCTGCAAGGGGTGCCAGAGACCCC	43160
QY	2999	ACCTGCTGTCCCAAGGGAACCCCAAGTGTAAATCTGTGGCTGTGAGTTTCACG	3058
Db	43161	ACCTGCTGTGTCCCAAGGGAACCCCAAGTGTAAATCTGTGGCTGTGAGTTTCACG	43220
QY	3059	TACAAGACCCCAAGAGTCCCTACTCCATCCCATCCAGTCCGCCCTTGCCGCCACACCC	3118
Db	43221	TACAAGACCCCAAGAGTCCCTACTCCATCCCATCCAGTCCGCCCTTGCCGCCACACCC	43280
QY	3119	CACCCCGCACTCCCGTCCACTTCTCTAGGCGCTGAAGGTCGCCAGCCCTGTGGGGTT	3178
Db	43281	CACCCCGCACTCCCGTCCACTTCTCTAGGCGCTGAAGGTCGCCAGCCCTGTGGGGGTT	43340
QY	3179	GCCTACCTCCAGTGTAGCGCCAGGTCCTAAGCCGGAAGTGCACCCCATCCCTGAACCTGA	3238
Db	43341	GCCTACCTCCAGTGTAGCGCCAGGTCCTAAGCCGGAAGTGCACCCCATCCCTGAACCTGA	43400
QY	3239	GAGCCAAAGGGCGGGGACACACGCGACCTCAGGCTGTACAGGCTTGTCTGGGCTCTAGTTTC	3298
Db	43401	GAGCCAAAGGGCGGGGACACACGCGACCTCAGGCTGTACAGGCTTGTCTGGGCTCTAGTTTC	43460
QY	3299	CCAGGAGACCTGGGCACTACTTCCCAACCCCCCATCCATCTCTGTGGGCCCTATCTT	3358
Db	43461	CCAGGAGACCTGGGCACTACTTCCCAACCCCCCATCCATCTCTGTGGGCCCTATCTT	43520
QY	3359	CCCTTATATGTGTGAAGAAGTTCTGTGGGGGGGGGGTGGTGAAGACAAAGTCTGT	3418
Db	43521	CCCTTATATGTGTGAAGAAGTTCTT - GGGGGGGGGGGTGGTGAAGACAAAGGTCGT	43578
QY	3419	CGGTCCTCGACGCGAGCTTGGCACAACTCTCTAAGATCTCCAGGTGGTGGCTGCCTT	3478
Db	43579	CGGTCCTCGACGCGAGCTTGGCACAACTCTCTAAGATCTCCAGGTGGTGGCTGCCTCT	43638
QY	3479	TCCAGACAGAGTAAGGCAATTGGGTGGGGACACATGGTGCACACAGTGGTTGGAGGGAC	3538
Db	43639	TCCAGACAGAGTAAGGCAATTGGGTGGGGACACATGGTGCACACAGTGGTTGGAGGGAC	43698
QY	3539	AGGGTCCTTGTCTCTCTGGGAGGCTGTGCTTTCTGTAGCACCTTGGTATTAAGTTTGGG	3598
Db	43699	AGGGTCCTTGTCTCTCTGGGAGGCTGTGCTTTCTGTAGCACCTTGGTATTAAGTTTGGG	43758
QY	3599	GGTGAAGTAAGTGGCTCTAAACCTCGAAAGAAGCAAGAAGCACAGGCTGTCTTGGGC	3658
Db	43759	GGTGAAGTAAGTGGCTCTAAACCTCGAAAGAAGCAAGAAGCACAGGCTGTCTTGGGC	43818
QY	3659	CTTTCATGAAGAAGTTCACAGACCCCTTTCCTGTAACTCACTTCCGCTCATCTGTGT	3718
Db	43819	CTTTCATGAAGAAGTTCACAGACCCCTTTCCTGTAACTCACTTCCGCTCATCTGTGT	43878
QY	3719	AGATTCCCTGGGACCAAGTGGCTCTGGGACTCAGATTTCTACAAATTAAATCAGGACA	3778
Db	43879	AGATTCCCTGGGACCAAGTGGCTCTGGGACTCAGATTTCTACAAATTAAATCAGGACA	43938
QY	3779	GTCTGAGACCTTGGACTCCGCTCGCTGATTTAACTACTTCTCTCTGGCTCTCATTTCTGT	3838
Db	43939	GTCTGAGACCTTGGACTCCGCTCGCTGATTTAACTACTTCTCTCTGGCTCTCATTTCTGT	43998
QY	3839	GTTTCATGTCTTACACATCTGAATGGTTTCTTTGTGTACCAATTCCTCCCTGACACTCTGG	3898
Db	43999	GTTTCATGTCTTACACATCTGAATGGTTTCTTTGTGTACCAATTCCTCCCTGACACTCTGG	44058
QY	3899	GAGGTCTGTATCTTGGCACATGTATCTGTGGATGTAACTGCAAGCCACACAGAGAGAGGG	3958
Db	44059	GAGGTCTGTATCTTGGCACATGTATCTGTGGATGTAACTGCAAGCCACACAGAGAGAGGG	44118
QY	3959	GGAGAGTACGAGAGCTGTGTCTTAAGGCCCTAATTAAGGCTGGACATCACCCCTTCTAGAA	4018
Db	44119	GGAGAGTACGAGAGCTGTGTCTTAAGGCCCTAATTAAGGCTGGACATCACCCCTTCTAGAA	44178
QY	4019	ATGGGCCCTCCATTTTTCGGTTACCATATGATTTATATATACAGATGGGACGTGAAMGC	4078

```
|||||
Db 44179 ATGGCCCTCCATTTTGGTTACCATGATCTATTTTATTCAGAGTGGGACGTAAAGC 44238
OY 4079 CAACCTGCCCCAGAGTTTGGGACTCTACTCAGACCAAGGTTATCTGTCAGAAATCCCC 4138
Db 44239 CAACCTGCCCCAGAGTTTGGGACTCTACTCAGACCAAGGTTATCTGTCAGAAATCCCC 44238
OY 4139 TGTCACTTGAAGTTGGGAGAACTCTGCTGGGGCTCCAGGCTCTGCTGAGAGAG 4198
Db 44239 TGTCACTTGAAGTTGGGAGAACTCTGCTGGGGCTCCAGGCTCTGCTGAGAGAG 44358
OY 4199 GTATCCCTTTGTATAGGGCATGACCTAGTCTATGTTGTTACTACATCTCTGTCAGTTAA 4258
Db 44359 GTATCCCTTTGTATAGGGCATGACCTAGTCTATGTTGTTACTACATCTCTGTCAGTTAA 44418
OY 4239 AGCTGGAACCTAAACCCAGCGGAGCCAGGATTCCTACAGTTGTACCCCAAGACAA 4318
Db 44419 AGCTGGAACCTAAACCCAGCGGAGCCAGGATTCCTACAGTTGTACCCCAAGACAA 44478
OY 4319 CAAGCACTAGATATGCAAGGATAGTGTGGGAGAAAGAACTTAACCCCCCAA 4378
Db 44479 CAAGCACTAGATATGCAAGGATAGTGTGGGAGAAAGAACTTAACCCCCCAA 44538
OY 4379 AGGCCACAGGTTCCGTTCCCTAGTTCACAAATGCCAGTATAGTCTAGCTATGAGC 4438
Db 44539 AGGCCACAGGTTCCGTTCCCTAGTTCACAAATGCCAGTATAGTCTAGCTATGAGC 44538
OY 4439 TGTGAGTTGGTAGCTACAGACATGAGTATGTCATGTTGTTAGTGTATTAATCTGAGC 4498
Db 44599 TGTGAGTTGGTAGCTACAGACATGAGTATGTCATGTTGTTAGTGTATTAATCTGAGC 44658
OY 4499 ACTTGGGAGGCTGAAAGCAGGAGATGCTATATGTTTATAGGCCACACCTGAGCTATAGAGC 4558
Db 44659 ACTTGGGAGGCTGAAAGCAGGAGATGCTATATGTTTATAGGCCACACCTGAGCTATAGAGC 44718
OY 4559 GAGACTTGTCTTTAAGAAAAAATGAAGCCACAGCAGTGGGAGACAGCCTTTAATCC 4618
Db 44719 GAGACTTGTCTTTAAGAAAAAATGAAGCCACAGCAGTGGGAGACAGCCTTTAATCC 44778
OY 4619 CAGCACTTGGAGGACAGACAGCAGATTTCTGAGTTCAAGGCCAGCCTGGTCTATAGA 4678
Db 44779 CAGCACTTGGAGGACAGACAGCAGATTTCTGAGTTCAAGGCCAGCCTGGTCTATAGA 44838
OY 4679 GTGAGTTCCAGGACAGCCAGGCTACACAGAAACCCCTGTTTGAATAACAGAAAC 4738
Db 44839 GTGAGTTCCAGGACAGCCAGGCTACACAGAAACCCCTGTTTGAATAACAGAAAC 44898
OY 4739 AAAACAAAACAAAACAAAACAAAACCCAAACCCAAACCCCTCATCTCATCTC 4798
Db 44899 AAAACAAAACAAAACAAAACAAAACCCAAACCCAAACCCCTCATCTCATCTC 44958
OY 4799 TCTAGGCTGTCTGTCTCTAGTGTAGAGTTTGGGACTTCAGACTTATATATAATAGG 4858
Db 44959 TCTAGGCTGTCTGTCTCTAGTGTGTAGAGTTTGGGACTTCAGACTTATATATAATAGG 45018
OY 4859 CCTTTTATCACTGCTGAGACGAGAAAGTTTCAGTCTGGGACACAGTGGGAACTCGA 4918
Db 45019 CCTTTTATCACTGCTGAGACGAGAAAGTTTTCAGTCTGGGACACAGTGGGAACTCGA 45078
OY 4919 GAAAGTACTCTTGTCCAGCCAAAATTTCTGGAGAGCTTCTGGAGAGAGTGTCCG 4978
Db 45079 GAAAGTACTCTTGTCCAGCCAAAATTTCTGGAGAGCTTCTGGAGAGAGTGTCCG 45138
OY 4979 A----- 4979
Db 45139 ATTGAGTGTGAAGAAATATGAAGAAAGTGGAAAGCCAAAGCAGAAAGCCTGGGTGA 45198
OY 4980 -----TCAGACTTACTGTCTTAGAAGCAGAAAGAGAGGTTGGAAGATGTTGGTG 5029
Db 45199 AAGCAGCAGTCACACTACTTCTTAGAAGCAGAAAGAGAGGTTGGAAGATGTTGGTG 45258
OY 5030 GACAGACAGTTGGAACAGAGAGCAGAGGAGGAGCATCCAAAGTTTGTGAACATGTAGC 5089
|||||

Db 45259 GACAGACAGTTGGAACAGAGAGCAGAGGAGGAGCATCCAAAGTTTGTGAACATGTAGC 45318
OY 5090 TGACTTTTGGTCTCTGGGTGACAAAGTGTCCCCAGGATAGGGCTGTAGAAAGGGGACC 5149
Db 45319 TGACTTTTGGTCTCTGGGTGACAAAGTGTCCCCAGGATAGGGCTGTAGAAAGGGGACC 45378
OY 5150 AGGGGTGAGCCAAATGAGTTCAAGTTGAGGAGACATCCAGCCCAAGGTCCTGCTGGCAA 5209
Db 45379 AGGGGTGAGCCAAATGAGTTCAAGTTGAGGAGACATCCAGCCCAAGGTCCTGCTGGCAA 45438
OY 5210 GCTTAAGAATGAGAGCCCTCTAACCCTCCGTAAGTTTATAGGGAACAGAGAGCTGAG 5269
Db 45439 GCTTAAGAATGAGAGCCCTCTAACCCTCCGTAAGTTTATAGGGAACAGAGAGCTGAG 45498
OY 5270 AGATCCTTCTAGGCTGAAGAGAGGTATCTGCTGTGACCAACATGCTAGAGCAGAAAC 5329
Db 45499 AGATCCTTCTAGGCTGAAGAGAGGTATCTGCTGTGACCAACATGCTAGAGCAGAAAC 45558
OY 5330 AGTTGGACAGTTTACCCTTCAGAAACAGCCATCCCTCTTGGCTCTAAGAGAGGCTGGCC 5389
Db 45559 AGTTGGACAGTTTACCCTTCAGAAACAGCCATCCCTCTTGGCTCTAAGAGAGGCTGGCC 45618
OY 5390 CCTTCTGTTTAAAGATCTTACTTCTTTCAGAGAGAGGACAGCAAGCCTTTGTCCTCC 5449
Db 45619 CCTTCTGTTTAAAGATCTTACTTCTTTCAGAGAGAGGACAGCAAGCCTTTGTCCTCC 45678
OY 5450 CTGTTGTCATTAACACCCCTGTGTGTATACATTAAGTTTATTTTACTGTCACTTGTCTCC 5509
Db 45679 CTGTTGTCATTAACACCCCTGTGTGTATACATTAAGTTTATTTTACTGTCACTTGTCTCC 45738
OY 5510 AGGACAGTCATCTGCTGAGCTCTGCTGCTACCTACACCAAGATATGGCCCATCTCTCC 5569
Db 45739 AGGACAGTCATCTGCTGAGCTCTGCTGCTACCTACACCAAGATATGGCCCATCTCTCC 45798
OY 5570 ACCCAGAGAGTGCAGAGAGAGCCTTAGAGAAAGGTTAACAGTAAACAGATGGCCAGA 5629
Db 45799 ACCCAGAGAGTGCAGAGAGAGCCTTAGAGAAAGGTTAACAGTAAACAGATGGCCAGA 45858
OY 5630 ATAAACAAAACCTACTCTCTCTTGTACCCAAATTTGTTTGTCTGACCCAGAGAGGAGTGTG 5689
Db 45859 ATAAACAAAACCTACTCTCTCTTGTACCCAAATTTGTTTGTCTGACCCAGAGAGGAGTGTG 45918
OY 5690 TGTGAGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 5749
Db 45919 TGTGAGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 45970
OY 5750 TGTGTGTGTCTTGGGGGACTTTTTCATGCTAAAGAAATATGATATTGGGCCCATGCC 5809
Db 45971 TGTGTGTGTCTTGGGGGACTTTTTCATGCTAAAGAAATATGATATTGGGCCCATGCC 46030
OY 5810 AACAGGGGATTTGGGAGACAGTACGCTTCTGCAACACAGTACGCTGCCCAAGATGATTT 5869
Db 46031 AACAGGGGATTTGGGAGACAGTACGCTTCTGCAACACAGTACGCTGCCCAAGATGATTT 46090
OY 5870 GGTGGCTGGAATCACCAAGGGGACGCTGATCAGTGTGGAGACAGATCACAAAGTTAAGC 5929
Db 46091 GGTGGCTGGAATCACCAAGGGGACGCTGATCAGTGTGGAGACAGATCACAAAGTTAAGC 46150
OY 5930 CACCCCTGTGGGGCTCAGAGAGAGGAGTTTACAAAGAGTTTAAAGGCCCAACCACTTTTATTC 5989
Db 46151 CACCCCTGTGGGGCTCAGAGAGAGGAGTTTACAAAGAGTTTAAAGGCCCAACCACTTTTATTC 46210
OY 5990 CAAGCATGACTCAAAATCAAAATGTCAGAGAGATTAAGCTGTGAGAGATGGGGCTGTAG 6049
Db 46211 CAAGCATGACTCAAAATCAAAATGTCAGAGAGATTAAGCTGTGAGAGATGGGGCTGTAG 46270
OY 6050 TGTGGGACACTGCACTTGTGCACTTATTTAGTCACTATGAGGCCAAGAGACATCAAGAGGGTG 6109
Db 46271 TGTGGGACACTGCACTTGTGCACTTATTTAGTCACTATGAGGCCAAGAGACATCAAGAGGGTG 46330
OY 6110 ACTGGGTCCTACTCAGCTTGTGAGCAGCAGTGTGAATTTGGGTGACCTCACTCCGTGATGG 6169
Db 46331 ACTGGGTCCTACTCAGCTTGTGAGCAGCAGTGTGAATTTGGGTGACCTCACTCCGTGATGG 46390
|||||
```



QY	6170	AGAGGGCTGAGCACCACCAAGTACAGATGTTCCCTGTCTCATATGCCAGATTTCTGGCC	6229
Db	46391	AGAGGGCTGAGCACCACCAAGTACAGATGTTCCCTGTCTCATATGCCAGATTTCTGGCC	46450
QY	6230	AGTTTTCAAAGACATTAAGACATCATCTCGTGGTGGAAACAAAGTATCCAAAGCCCTAAGGCC	6289
Db	46451	AGTTTTCAAAGACATTAAGACATCATCTCGTGGTGGAAACAAAGTATCCAAAGCCCTAAGGCC	46510
QY	6290	CATTTTGGTCTAATTAAATCAGAACCCCTGGGGATGACAGGCTTGAGAGCAGAGACTTT	6349
Db	46511	CATTTTGGTCTAATTAAATCAGAACCCCTGGGGATGAGAGGCTTGAGAGCAGAGACTTT	46570
QY	6350	TTTAAAGACTCCCAAGSTGATTTCTGATTCAGACAGCTGGAAACAAACAGCTACAGTTCCAA	6409
Db	46571	TTTAAAGACTCCCAAGSTGATTTCTGATTCAGACAGCTGGAAACAAACAGCTACAGTTCCAA	46630
QY	6410	CAGAAAGAGGCAAGACTAAGGCAACCTTGGATGGGAGCCCTTCTCCAGGCACTAATAT	6469
Db	46631	CAGAAAGAGGCAAGACTAAGGCAACCTTGGATGGGAGCCCTTCTCCAGGCACTAATAT	46690
QY	6470	GGAGGCTGGTTAGCAGTGTGTGGACGCTTCTCTGCTCATATATAGCTATTCATCCACT	6529
Db	46691	GGAGGCTGGTTAGCAGTGTGTGGACGCTTCTCTGCTCATATATAGCTATTCATCCACT	46750
QY	6530	CATCCATCATTAACACCCACCCATCCATTATATACACCCATCCCTCCATCATCATCTATTC	6589
Db	46751	CATCCATCATTAACACCCACCCATCCATTATATACACCCATCCCTCCATCATCATCTATTC	46810
QY	6590	CAGCTAACCCACCCACGCACTCATCCAAACCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCT	6649
Db	46811	CAGCTAACCCACCCACGCACTCATCCAAACCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCT	46870
QY	6650	CTTCACTCATTAATTATATCCAAACAGAAAGCTGTATGACTAAATATGGGAGATTTAT	6709
Db	46871	CTTCACTCATTAATTATATCCAAACAGAAAGCTGTATGACTAAATATGGGAGATTTAT	46930
QY	6710	TAAATTTTGAAGAGCTCTGTTGATGACTGATTTGTCATGTATGTGGACAGGTATATACA	6769
Db	46931	TAAATTTTGAAGAGCTCTGTTGATGACTGATTTGTCATGTATGTGGACAGGTATATACA	46990
QY	6770	CAGCACACGCTGTGGCAATTCGGAGAAAGTTTGGGATGTTTCTCTCTCTCCACCGTGTG	6829
Db	46991	CAGCACACGCTGTGGCAATTCGGAGAAAGTTTGGGATGTTTCTCTCTCTCCACCGTGTG	47050
QY	6830	GCTTCTGGGGAATGAACATCAAAATATCGGGGCTGTGGCAAGTGTCTTACACCCAGGCA	6889
Db	47051	GCTTCTGGGGAATGAACATCAAAATATCGGGGCTGTGGCAAGTGTCTTACACCCAGGCA	47110
QY	6890	TTTTGCTGACACATCATTAATTTTGAAGAACATCTTATATGATGATCCAGGCTGGCCTCAAGC	6949
Db	47111	TTTTGCTGACACATCATTAATTTTGAAGAACATCTTATATGATGATCCAGGCTGGCCTCAAGC	47170
QY	6950	TTGCTATGTCGGCAGGATGACCTTAACTCTGCTCTCCAGGCTCCACCCGATGTGTA	7009
Db	47171	TTGCTATGTCGGCAGGATGACCTTAACTCTGCTCTCCAGGCTCCACCCGATGTGTA	47230
QY	7010	GGTTTACACGCTGTCAAC-----	7027
Db	47231	GGTTTACACGCTGTCAAC-----	47290
QY	7028	-----	7060
Db	47291	ACAAATCTGCGCAACAGCTGGGCAATGTGTGTAATGCTTAAATCTCCACACACTCTGTGGG	47350
QY	7061	GGGGGGGGGGAGGGAGATCCCTGATTTGGAGGCCAGTTTGGTCTTACAGAGATTTAGAT	7120
Db	47351	GGGGGGGGGGAGGGAGATCCCTGATTTGGAGGCCAGTTTGGTCTTACAGAGATTTAGAT	47410
QY	7121	ACCTGGGGCTATACAGGGAACCTTATCCCAAAACAAACAAACAAACAAATATTT	7180
Db	47411	ACCTGGGGCTATATACAGGGAACCTTATCCCAAAACAAACAAACAAACAAATATTT	47470

[illegible]



```

Db 48551 TATGAGGATATTTTGTGGAGTTTTCACCTTCCCTGTGGGCTCCGCATTAACCTCAG 48610
|
|
|
Qy 8313 CTCCTCGGGCTAGTAGAGCAATGCCCTTCACTGATGAGCAATCTCGCTCCCTGCTGCCA 8372
|
|
|
Db 48611 CTCCTCGGGCTAGTAGAGCAATGCCCTTCACTGATGAGCAATCTCGCTCCCTGCTGCCA 48670
|
|
|
Qy 8373 CCTCCTCTTATTTTCCCGATGGGCACTAGCACTGACCTGAGGCTTAAAGCTCACCAGTCA 8432
|
|
|
Db 48671 CCTCCTCTTATTTTCCCGATGGGCACTAGCACTGACCTGAGGCTTAAAGCTCACCAGTCA 48730
|
|
|
Qy 8433 TCCAGAGTGGCTAGCCAGGAGGAGTACAGGATATGCTGGCTTCCCTCCACAGTGTAG 8492
|
|
|
Db 48731 TCCAGAGTGGCTAGCCAGGAGGAGTACAGGATATGCTGGCTTCCCTCCACAGTGTAG 48790
|
|
|
Qy 8493 AATTACAGGCTACATCACTGCTGGAAATTTTAACTGAACTCTGAGATAGACGACG 8552
|
|
|
Db 48791 AATTACAGGCTACATCACTGCTGGAAATTTTAACTGAACTCTGAGATAGACGACG 48850
|
|
|
Qy 8553 CACTCTACCAATGAGAGGCTCTTTTGTGTGGTTGGTTGCTTCCCTGCAATAGATCAGG 8612
|
|
|
Db 48851 CACTCTACCAATGAGAGGCTCTTTTGTGTGGTTGGTTGCTTCCCTGCAATAGATCAGG 48910
|
|
|
Qy 8613 CAGTCTGAATAGTGTAGCTTGGGCTACATACATCTTGTCTCAAAAAGCCTATAGAGT 8672
|
|
|
Db 48911 CAGTCTGAATAGTGTAGCTTGGGCTACATACATCTTGTCTCAAAAAGCCTATAGAGT 48970
|
|
|
Qy 8673 AGGAGGCTGAGGCTTAAAGAGAGCCTTAAAGCCGCTGTATAGCACACAGATAGCCTG 8732
|
|
|
Db 48971 AGGAGGCTGAGGCTTAAAGAGAGCCTTAAAGCCGCTGTATAGCACACAGATAGCCTG 49030
|
|
|
Qy 8733 CACTATATAGAGAGACCTTGTTCAAAACATGGAGGAGGCTATGTTTAAAGGCTGG 8792
|
|
|
Db 49031 CACTATATAGAGAGACCTTGTTCAAAACATGGAGGAGGCTATGTTTAAAGGCTGG 49090
|
|
|
Qy 8793 GCTGTGTAAAGGAGCACTAAGGAGGCAATGTAGACATTTGACTAAGAAGATCATCATC 8852
|
|
|
Db 49091 GCTGTGTAAAGGAGCACTAAGGAGGCAATGTAGACATTTGACTAAGAAGATCATCATC 49150
|
|
|
Qy 8853 AAAGCCGGGTGGGCAAGGCTAAGGTTGACATACAGTGTCAAGAGCCCATAGAGGCCA 8912
|
|
|
Db 49151 AAAGCCGGGTGGGCAAGGCTAAGGTTGACATACAGTGTCAAGAGCCCATAGAGGCCA 49210
|
|
|
Qy 8913 GTTTCCTCTTCTTCCTCGGGCTCAAGCTGAGCTGAGGCACTGCTCACAATGCTT 8972
|
|
|
Db 49211 GTTTCCTCTTCTTCCTCGGGCTCAAGCTGAGCTGAGGCACTGCTCACAATGCTT 49270
|
|
|
Qy 8973 CTCCTTAGGCTGCTCCACCATG 8995
|
|
|
Db 49271 CTCCTTAGGCTGCTCCACCATG 49293
|
|
|
RESULT 3
AC101018 149109 bp DNA linear HTG 21-AUG-2002
LOCUS Rattus norvegicus clone RP32-328P7, *** SEQUENCING IN PROGRESS ***,
DEFINITION 9 unordered pieces.
ACCESSION AC101018
VERSION AC101018.2 GI:22381320
KEYWORDS HTG; HTGS_PHASE1; HTGS_FUZZTOP; HTGS_ACTIVEFIN.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 149109)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Rattus norvegicus, clone RP32-328P7
JOURNAL Unpublished
2 (bases 1 to 149109)
REFERENCE Birren,B., Linton,D., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
AUTHORS Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Camarata,J., Chang,J., Chazaro,B., Boukhalter,B.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
3 (bases 1 to 149109)
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
Direct Submission
Zainoun,J., Zember,L., and Zody,M.
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17059792.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14794
Center clone name: 328_P_7
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 46829: contig of 46829 bp in length
* 46830 46929: gap of 100 bp
* 46930 49621: contig of 2692 bp in length
* 49622 49721: gap of 100 bp
* 49722 70009: contig of 20288 bp in length
* 70010 70109: gap of 100 bp
* 70110 77196: contig of 7087 bp in length
* 77197 77296: gap of 100 bp

```





Dh 100265 AGATTTTACAAATTAATCATGACAGACATCTGAGAAAGCTGGAAATTAAGTGTCTCTCT 100324  
Qy 3799 TGCCGTATATTACTACTTCTCTGCGGCTCATTTCTGTTGATGCTTTACACATCTG 3858  
Dh 100325 TGGCAACATTTCTTCTTCTCTCTGAGGTTCAATTTCTTGTCTATGTCTTGACCTTCTG 100384  
Qy 3859 AAATGCG---TTTCTTTGTGTCAACATTTCCCTGACACTCTCTGGAGGCTGTATCTTTGG 3914  
Dh 100385 AAATGCTTCTTCTTGTGTGTGACACATTTCCCTGACACTCTGTGAGAGGTTGGGCTTGG 100444  
Qy 3915 CAC---ATGATATCTGGAGATATAGCTGCACCCACACAGA-GAGAGGGGAGAGTACG 3969  
Dh 100445 CCCACCATGATCTCCGTGATGTAAAGCTGCAGTCACACAGAGGGGAAAGAGAGCTGG 100504  
Qy 3970 AG---CTGTGTCTAGAGCCCTATTAGCCGTGACATCACCCCTTTCTTAAGAAATGGCCC 4025  
Dh 100505 AGGCTACAGGCTCTAGGCCCTCTCTGCTGTGATCATCTCTGCTCCCTTGAATATGGCCC 100564  
Qy 4026 CTCATTTTTCGGTTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGCCAAACT 4085  
Dh 100565 CTCTGCTTTCAGTCTCCGTGTCTATTT---TATCAGAGCTGACAGTGAAGCAAGCCACT 100622  
Qy 4086 GCCCAGAAAGTTTGGGACTCACTGACACCAAGGTTATCTGCTCAGAAATCCCTCTGACT 4145  
Dh 100623 GCCCAGAAATTTTGGGACTCACTTAACCAAGTCTATGTCTCAGAGCTCTTTTGTCACT 100682  
Qy 4146 T-GAGGTGGAGAAATCGCTCTGCGGGGCTCCAGGCTTTGGTTTACAGAGGATATCC 4204  
Dh 100683 TGGAGGTTGGAGATATCTCTCTCCAGGGGCTCCAGGCTCTGTTTACAGAGGAGGCTCC 100742  
Qy 4205 TTTGTATAG---GCATGACCTAGTCTATGTGTGTACTACATCTCTCCAGTTAAAGC 4261  
Dh 100743 TTTGTATAGGAACCTGACCTAACCCATGATGTGTACTACACCTCTCTCCAGTTAAAGC 100802  
Qy 4262 TGGAACTAAACCACGCGCAGCGCCAGAGTTCTCTACAGTTGTACCCCAAGAACAA 4321  
Dh 100803 TGGGACTAAACCACCGTGCAGCACCAGAGATTCCTATGTGTGACCTCCAGAACAG 100862  
Qy 4322 GACAGTATATGCAAGATAGTATAGTGGGAGAGAAAGAACTTAA-ACCCCCCAAG 4380  
Dh 100863 GACAGTATATGCAAGAGAGAGAGTATGCTGAGAGAGAGAACTTAAAGTCTTAAGGCCAT 100922  
Qy 4381 GCCCAGAGTTCCGTTCCCTAGTTCAACAATGCCAGTATGAGTGTAG-----CTACT 4432  
Dh 100923 AGGTCTCTCTCCGTTCCCTAGTTCTGAATGTGATGTGATGCCAATGGATTTCTACT 100982  
Qy 4433 ATGGGCTGTGAGTTGTGATGATCAACAGCATGATGA-ITGTCATGTGTGTACTGTATTA 4491  
Dh 100983 ATGGGCTGTGAACTGTGATGATCAAGTGTGATGATGATGATGATGATGATTA 101042  
Qy 4492 TCTGAG-CACTTGGAGGCTGAAGCAGAGGATTTGATATGTTTGAAGCCAGCTGAGC 4550  
Dh 101043 TCTGAGCACTTGAAGGCGGAGCATG-----TTTCTATGTTTGAAGCCAGCTGAGC 101097  
Qy 4551 TATGAGCGAGACTTTTCTTTTAAAGAAATGAAGCCAGCAGTGTGTGACAGGCC 4610  
Dh 101098 TATGAGGAGAGAGCTGTCTTTTAAAGAAAGAAAGCAACACG----- 101144  
Qy 4611 TTTTATCCCAAGCACTTGGAGGAGAGAAAGCAAGATTTTGAAGTCAAGGCCAGCTGG 4670  
Dh 101145 ----- 101144  
Qy 4671 TCTATAGATGATTCAGAGACAGCCAGGCTTCACAGAGAAACCTGTTTGAAAAAC 4730  
Dh 101145 ----- 101144  
Qy 4731 AGAAAAACAAACAAACAAACAAACAAACCAACCAACCAACCTCTCATCT 4790  
Dh 101145 -----ACCTTAAACAAACAAACAAACAAACCAACCAACCAACCAAA 101189  
Qy 4791 CTCTATCTCTAGGCTGT 4850  
Dh 101190 TTTCTATCTGTAGGTTGT 101249

Qy 4851 TAAATAGCCCTTTTATACATGTCAGACAGAGAAAGTTTCAAGTCTGAGACAGATG 4910  
Dh 101250 TAGGCCCTTTTATCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 101309  
Qy 4911 GACCTTGAGAAAGTACTCTTGTCCAGCCCAAAATTTCTGGGAAGCTTCTGTGAGGAAT 4970  
Dh 101310 GTCCCTGAGAAAGACTCTTGCACAGCTTAGAAATTTCTGGGAAGGCTTCTGTGAGGAAT 101369  
Qy 4971 GTGTCT-----CCGATCA 4982  
Dh 101370 GTCCGATTTGATGCTGAGAAATAATGAAAGTGGGAAAGCCAGACAGACAGCTGTAG 101429  
Qy 4983 GACTACTGTCTGAAGAGCAGAGAGAGGTTGGAAGATGT-----T 5025  
Dh 101430 AACTACTGTCTGAAGAGCAGAGAGAGGCTTGAAGAGTGTCTGTACGACAGAGCGAT 101489  
Qy 5026 GTGTGACAGACAGTTGGAACAGAGACAGG---AGGGGAGGCAATCAAGATTTTGAA 5081  
Dh 101490 GTGTGACAGACAGTTGGAACAGAGGACTTGGGACAGGGGAAACCGTCAAGATTTCTGA 101549  
Qy 5082 CATGTACTGACATTTTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5141  
Dh 101550 CACATAGCTGTGCTTGT 101597  
Qy 5142 AGGGAGACAGGGGTGAGCAGCAATGATTTCAAGTTGAGGAGACATCCAGCCAGGGTCTT 5201  
Dh 101598 TGGGAGACAGGGGTGAGTCACTAGTCAAGTTGAGGAGACATATCTTCCACAGGCTCTT 101657  
Qy 5202 GCTGTGACAGCTTAAAGATGAGAGCCCTTAAACCTCCCTGAAGTTTGGGAGACAGAG 5261  
Dh 101658 GCTGTGACAGCTTAAAGATGAGAG--CTTCAACCTCTTGAAGTTTGGGAGACAGTGG 101716  
Qy 5262 AGCTGAGAGATCTCTGTAGGTGAAGAGAGATGT--GCTGTGACCAATGTGTGTAGG 5320  
Dh 101717 AGCTGAGAGATCTCTGTAGGTGAAGAGAGATGTCTCTGTGACCGATGTGTGTGGA 101776  
Qy 5321 AGCAGAGCA-----GTTGACAGCTTACCCCTGACAGACAGCCATC 5362  
Dh 101777 AGCAGAGCACTTGAGATACCCAGAGGAGTTGGCCGCGCTGTCTGACAGCCAGCATC 101836  
Qy 5363 CCCTCTTGGCTTAAGAGGCTGTGGCCCTTCTGTAAAGATCTTATTTCTTCTGAGA 5422  
Dh 101837 CCACCTTGGCCCTTAAGAGGCTGTGGCCCTTCTGTAAAGATCTTCTGAGA 101895  
Qy 5423 GAGAGGAGAGAGCTTGT 5482  
Dh 101896 GAGAGGAGAGAGCTTGT 101954  
Qy 5483 TAGTTTATTTTACGTGATTTGT 5542  
Dh 101955 TAGTTTATTTTGT 102014  
Qy 5543 CTCACCAAGTATGCCCCCATCTTCCACCCAGAGAGTGCAGAGAGAGAGCTTGTAGAA 5602  
Dh 102015 CTCACCAAGTATGCCCCCATCTTCCACCCAG--AGTGTCAAAAGAGAGAGCTTGTAGAA 102073  
Qy 5603 AGGTTAAGTATTAAGATGTGCGCAGATTAACAAACAACTACTATCTTGTATCCCAA 5662  
Dh 102074 AGGTTAAGTATTAAGAGAGAGAGAGAGATTAACAAACAACTACTATCTTGTATCCCAA 102133  
Qy 5663 TTTGTTTGTGGAACAGAGAGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5722  
Dh 102134 -----TTGCTGAACAGGA----- 102147  
Qy 5723 GT 5782  
Dh 102148 -----GTGAGT 102189  
Qy 5783 AGAATATCTATATTTGGCCGCTATGCCAACAGGGGTATTGGGAGAGCT-----AG 5833  
Dh 102190 GGAATATCTATATTTGGGCGCCTATGCCAATAGGGGTATTCAGGTAGAGTACAGGCTGTAG 102249

```
OY 5834 GCTTCGCAACACAGTAGCTGCCAAGATGATTGGGCGCTGATCCACAGGGGA 5893
|||||
Db 102250 GCTTTCGCAACACAGCAAGCTGCCCAAGATGATTGGTGGCTTGATCACCAGGGGA 102309
OY 5894 GCGTCGACAGAGTGGAGCAACATCAAGATTAAAGCCACCCGTGGGGCTGAGAAGGG 5953
|||||
Db 102310 CGCTGATCAGAGTGAAGCAAGAA---CACAAGATTAAGCCACCCGTGAGGGCTGAGA-----G 102361
OY 5954 AGTTTACAAGAGTTAAAGGCCAGCCATTATTATTCCAAGACATGACTCAAAATCAAGT 6013
|||||
Db 102362 AGTTTACAAGAGTTAAAGGCCAAGCCATTATTATTCCAAGACATGACTCAAAATCAAGT 102421
OY 6014 GCAAGGAGAGATTAGCTGGAGATGGGGCTGCTGAGTGGGACACTGACCTTGCACTT 6073
|||||
Db 102422 GCAAGGAGAGATTAGCTGGAGATGGGGCTGCTGAGTGGGATGACTGACTTGCCCTT 102481
OY 6074 ATTACTACTAGGCCAAGAGACGTACAGAGAGGTGACTGGGTCTTACTGACCTTGGAGC 6133
|||||
Db 102482 GTTACTACTAGGCCAAGAGACGTACAGAGAGGTGACTGACTGCTACTGACCTGGGGC 102541
OY 6134 AGGCACGTGGAGAAATGGGTGACCTTCATCTGATGGAGAGGGCTAGACCAACAGGTAC 6193
|||||
Db 102542 AGGCACATAGAGAAATGGGTGG-----TGAGAGGGGTGAGATGACCAAGGTGC 102589
OY 6194 AAGTGTTC---TGTGCTCATGCCAGAGATTCTGGCCAGTTTCAAGGACTAAGGC 6250
|||||
Db 102590 AAGTGTTCCTCGATGTGCTGCTGCTGCTGGATTTCCAGCCAGTTTCAAGAGCCAGGTG 102649
OY 6251 TCATCTGTGTGGAACAAGATATCCAAGCCCTAAGCCCATTTTGTCTAATTAAATCA 6310
|||||
Db 102650 TCATCTGTGTGGAATTAAGATATCCAAGCCCTAAGCCCTATTGTGCTAATTAAATCA 102709
OY 6311 GAACCCCTGGGGATGACAGCTCTGAGCAGCAGAGACTTTTAAAAAGCTCCAGGTGAT 6370
|||||
Db 102710 GAACCTCTGGGGATGACAGCTCTGAGCAGCAG---TTTTTAAAAAGCTCCAGCGACT 102765
OY 6371 CTGATCAGC--AGCTGGAACAACACAGCTACAGG--TTCAAAACGAAGAAGAGCAAGCT 6426
|||||
Db 102766 CTGATCAGCAGAAATAGAAATAAAAACAGCTACAGGGTTCAACAGAAAAAGAGCAAGCT 102825
OY 6427 AGGGAAGCTTGGGATGGGAGCCCTTTCAGAGCCAGTAGATGGAGCTGGTTAGCAGT 6486
|||||
Db 102826 AGGGAAGCTTGGGAT--GGAAGCTTCTTTCAGCCAGTAGCAAGGCCAACCAANNNN 102883
OY 6487 GGTGGACACTCTCTCTGCTCTCATATAGCTATCCTACCTACTCATCATACACCC 6546
Db 102884 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 102943
OY 6547 ACCATCCATTATGACCCCATCTCTCCATCCATCTATCCAGCTAACCCACCCAGC 6606
Db 102944 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 102988
OY 6607 ATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6666
|||||
Db 102989 ACCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 103048
OY 6667 TCCACAAGAGACTGCTGTTGACTAAATGTTGGAGATTTAATTAATTTTGAAGCTCT 6726
Db 103049 CGTAACAGAGACTGTTGTTGACTAAATGAGAGATTTAATTAATTTAATTTAA--AAA 103106
OY 6727 GTTGATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6786
|||||
Db 103107 ATTATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 103166
OY 6787 TCGGAGAAAGTTTGGGTGTTGTTTCTCTCCACCGTGTGGTTC--TGGGATTTGAA 6845
|||||
Db 103167 TCGAGGAAAGTTTGGGTGTTGTTTCTCTCCATGTTGGGTCTCTTCGAGACTTAC 103226
OY 6846 CTCAAAATATCGGCTGTTGGGCAAGTGTCTTTA--CCACGAGCCATTTTGTGACATCAT 6904
|||||
Db 103227 CTCAAAATATCGACTGTGTGGCAAGTGTCTTTATTCATTTGACCCATTTTGTGCAAA---C 103283
OY 6905 ATTATTTATGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6964
|||||
Db 103284 CTAAATTTATGAAAGTGTCTTCTGTATACCAAGCTGGGCTCAGGCTTGCTATCATATGCCAC 103343
OY 6965 GGATGACCTTTTAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7024
|||||
Db 103344 ACATGACCTTTTAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 103403
OY 7025 A----- 7025
Db 103404 ACCGAGTGAATGCTGGGATTTATACGTATACGTGGGACAGCAAAATGTCTCCCTGCC 103463
OY 7026 -----ACTGGTAATGCTTTTATCCACAGCTC 7054
Db 103464 TTGTAGAAACATCTGGGCAACAGCTGGGCAATAGTGCTGATGCCCTTTGATCCACAGCTC 103523
OY 7055 TGTGGGGGGGGGGGGGAGGGGAGATCCCTGAGTTTGGAGGCCAGTTTGGTCTACAGATTT 7114
Db 103524 T----GAGGCAAGAGAGAGGAGATCCCTGAGTGGAGGCCAGTTTGGTCTACAGATTT 103579
OY 7115 CAGGATACCTGGGCTATACAGGGAACCCCTATCCCAACAACAA-----A 7161
Db 103580 CAGGATACCTGGGCTATACAGGGAACCCCTATCCCAACAACAAATATTCCTTAATC 103639
OY 7162 CAACAAACAAAAATATTCGTGCAATATCACAGAGATTAGAGATATTAGTAGGTA 7221
Db 103640 ACAGAGATTAGAGATGTTGGTGGGGCTGTGTAGGAGAGAGCATGCTTTCTTTGTATTA 103699
OY 7222 GTAGGGCTGTGTAGGAGAGATCATGCTCTTCTTTGTATTAATTAATTAAGTA----- 7274
Db 103700 TTATCTATCTGGATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 103759
OY 7275 ----- 7274
Db 103760 GGTGCTTTGACCCCTAAATTCCTCTATTCTGTGTGACCTCTTCTCAGTGTAGGGTTT 103819
OY 7275 -----CTCACAAGATGATTAATCTATCTATCTATCTATCTATCTATCTATCT 7313
Db 103820 ACAGGATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 103879
OY 7314 ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7340
Db 103880 ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 103939
OY 7341 TACCTACTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7387
Db 103940 TATCAAACTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 103999
OY 7388 GACTCTGAATGCTCTATTTCTGGGTCAACCTTCAACCCCTAGTGTGGTTTACACA 7447
Db 104000 ATCAAACTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 104059
OY 7448 CCCAGACTTAT-----T 7461
Db 104060 AACTATCAATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 104119
OY 7462 TTATTTTGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7521
Db 104120 CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 104179
OY 7522 TTAAGCAAGCTCTGCGCAGAGCTGAGCTGACGCTCCAGCTATTTGTTTCAGTACTCT 7581
Db 104180 CTAAACAGCTCTCTGTCACAGAGCTCCAGCCGAGCTCCAGCTTTG--TCAGGTGATCT 104238
OY 7582 GTGACAGTTGTCATATTTGCAAGCGCTATGTAAGCTCTCTCCACCTTCCAGCTT 7641
Db 104239 GTGGGCTTTGACATATTTCAAAATGATGATGATGATGATGATGATGATGATGATGATGAT 104298
OY 7642 TCTGGTCAATCCAGTGGGCGGCAACCTGCTGCTACAGAGGCGCTGTTCCCTG----- 7695
Db 104299 TCTGGTCAATCCAGT-----GGGCAAGCTGTGCTTACAGAGTACTGTGTCCCGTCCCC 104354
OY 7696 -----TCTTCAGACTTATATTTGCTGCTGCAAGCTTCAATGTA 7738
|||||
```



REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	COMMENT
Unpublished 2 (bases 1 to 159492) Worley, K. C. Direct Submission Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 159492) Worley, K. C. Direct Submission Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 12, 2002 this sequence version replaced gi:19482237.	Unpublished 2 (bases 1 to 159492) Worley, K. C. Direct Submission Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 159492) Worley, K. C. Direct Submission Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 12, 2002 this sequence version replaced gi:19482237.	
Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a> ----- Project Information Center project name: GQDN Center clone name: CH230-286017 ----- Summary Statistics Sequencing vector: Plasmid: Chemistry: Dye-terminator Big Dye: 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 124398 bases at least Q40 Consensus quality: 128059 bases at least Q30 Consensus quality: 131015 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html</a> ) * NOTE: This is a 'working draft' sequence. It currently * consists of 46 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a> ----- Project Information Center project name: GQDN Center clone name: CH230-286017 ----- Summary Statistics Sequencing vector: Plasmid: Chemistry: Dye-terminator Big Dye: 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 124398 bases at least Q40 Consensus quality: 128059 bases at least Q30 Consensus quality: 131015 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html</a> ) * NOTE: This is a 'working draft' sequence. It currently * consists of 46 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	
1 1015 1014: contig of 1014 bp in length 1115 1114: gap of unknown length 1135 2395: contig of 1281 bp in length 2396 2495: gap of unknown length 2496 3600: contig of 1105 bp in length 3601 3700: gap of unknown length 3701 5021: contig of 1321 bp in length 5022 5121: gap of unknown length 5122 6727: contig of 1606 bp in length 6728 6828: gap of unknown length 6828 8708: contig of 1781 bp in length 8609 8709 10003: contig of 1295 bp in length 10004 10103: gap of unknown length 10104 11211: contig of 1108 bp in length 11212 11311: gap of unknown length 11312 12518: contig of 1207 bp in length 12519 12618: gap of unknown length 12619 13843: contig of 1225 bp in length 13844 13943: gap of unknown length 13944 15404: contig of 1461 bp in length 15405 15505: gap of unknown length 15505 17647: contig of 2143 bp in length 17648 17747: gap of unknown length 17748 19706: contig of 1959 bp in length 19707 19806: gap of unknown length 19807 21576: contig of 1770 bp in length 21577 21676: gap of unknown length 21677 22906: contig of 1230 bp in length 22907 23007: gap of unknown length 23007 25086: contig of 2080 bp in length 25087 25186: gap of unknown length 25187 26832: contig of 1646 bp in length	1 1015 1014: contig of 1014 bp in length 1115 1114: gap of unknown length 1135 2395: contig of 1281 bp in length 2396 2495: gap of unknown length 2496 3600: contig of 1105 bp in length 3601 3700: gap of unknown length 3701 5021: contig of 1321 bp in length 5022 5121: gap of unknown length 5122 6727: contig of 1606 bp in length 6728 6828: gap of unknown length 6828 8708: contig of 1781 bp in length 8609 8709 10003: contig of 1295 bp in length 10004 10103: gap of unknown length 10104 11211: contig of 1108 bp in length 11212 11311: gap of unknown length 11312 12518: contig of 1207 bp in length 12519 12618: gap of unknown length 12619 13843: contig of 1225 bp in length 13844 13943: gap of unknown length 13944 15404: contig of 1461 bp in length 15405 15505: gap of unknown length 15505 17647: contig of 2143 bp in length 17648 17747: gap of unknown length 17748 19706: contig of 1959 bp in length 19707 19806: gap of unknown length 19807 21576: contig of 1770 bp in length 21577 21676: gap of unknown length 21677 22906: contig of 1230 bp in length 22907 23007: gap of unknown length 23007 25086: contig of 2080 bp in length 25087 25186: gap of unknown length 25187 26832: contig of 1646 bp in length	

DB	Accession	Score	DB 2	Length	159492
DB	48550	10.4%	Score 938.6	DB 2	Length 159492
			Best Local Similarity 63.7%		Pred. No. 1.2e-263
			Matches 1950	Conservative	0: Mismatches 664; Indels 447; Gaps 25;
DB	48550	10.4%	Score 938.6	DB 2	Length 159492
			Best Local Similarity 63.7%		Pred. No. 1.2e-263
			Matches 1950	Conservative	0: Mismatches 664; Indels 447; Gaps 25;
DB	48550	10.4%	Score 938.6	DB 2	Length 159492
			Best Local Similarity 63.7%		Pred. No. 1.2e-263
			Matches 1950	Conservative	0: Mismatches 664; Indels 447; Gaps 25;



[illegible][illegible]



[illegible]

REFERENCE	AUTHORS
1 (bases 1 to 159492)	Mizuy, D.M., Adams, C., Adio-Ogunola, B., Allison, F.R., Allen, C., Aisbrook, S.L., Ameralunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbada, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gottell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteberg, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozard, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Minner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwkw, S., Oguni, M., Okwunodu, G., Orangunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherger, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 159492)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 159492)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jul 12, 2002 this sequence version replaced gi:19482237.
	----- Genome Center
	Center: Baylor College of Medicine
	Center code: BCM
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>
	Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a>
	----- Project Information
	Center project name: GQDN
	Center clone name: CH230-286017
	----- Summary Statistics
	Sequencing vector: Plasmid;
	Chemistry: Dye-terminator Big Dye; 100% of reads
	Assembly program: Phrap; version 0.990329
	Consensus quality: 124398 bases at least Q40
	Consensus quality: 128059 bases at least Q30
	Consensus quality: 131015 bases at least Q20
	-----
	* NOTE: Estimated insert size may differ from sequence length



Db	28031	TGCTTTTCGGTTTCCTCCGGGTGACAGTGTGCTTCCACGGG-----ATGGGGACC	27988
Qy	5150	AGGGGTGAGCCATGAGTTCAAGTTGAGGGACATCCAGCCAGCGTCTTGTGCGCA	5209
Db	27983	AGGGGTGAGTCAAGTGAATTCAAGTTGAGGAACATATCTAACCCAGGGTCTTGTGGCAG	27924
Qy	5210	GCTAAGAATGAGACCCCTTAACCTCTCCCTGAAGTTAAGGGACACAGAGAGCTGAGG	5269
Db	27923	ACTGAAGATGAGAC-CCCTCCAAACCTCCCTTGAACTTTAAGGGGAGACTGGAGCTGAGG	27865
Qy	5270	AGATCCCTTATAGGGGGAAGAGGTATCT-GCCTGACCAACATGGCGTAGAGAGGAAG	5328
Db	27864	ATGTCTCTTGTGAGTGAAGGGAAGGTGTCTCCCTCTGACCGCATGAGGTGGAAGAGAGG	27805
Qy	5329	CA-----GTTGGACAGATTAACCTCTCAGAACCGACATCCCTCTG	5370
Db	27804	CACATTGAGGATATCCCAAGGGGTGTGGCCCGCGCTCTCAGAGCCAGCATCCACCTTG	27745
Qy	5371	GCTTAAGGAGGCTGGCCCTTCTGTGTTAAGAACTTACTTTTCTTCAGAGAGGCA	5430
Db	27744	GCCCTAAGAGAGCTGTGCCCTTGTGGTTAGATATCTAC-TTTCCTCAGAGAGAGA	27688
Qy	5431	GCAACCTTTTGTCCTCCCTCTGTGTGTAATAAACCCCTGTGTGTAATAGTTAT	5490
Db	27685	GCAACCTTTTGT-CCCGCGCAACAGTCAATTAACATCCCTGTGTCTAACATTAGTTAT	27627
Qy	5491	TTTACTGTACAGTTTCTCTCAGACAGTCCATGTGGTAGACCTCTGCTTACTACCCA	5550
Db	27626	TTTGCTGTACAGTTTCTCTCAGACAGTCCATGTGGTAGACCTCTGCTTACTACCCA	27567
Qy	5551	GGTATGGCCACATTTCTCACCACAGAAAGTGCAGAAAGAGCCCTTAGAAGAAAGGTAAC	5610
Db	27566	GCTATGGCCACCTTCTCACCACAG-AGTGTCAAAAGAGACGTTAGAGAAAGAGTAAC	27508
Qy	5611	AGTAAACAAGATGGCCAGATTAACCAAAACTACTATCTTTGTACCAATTGCTTTT	5670
Db	27507	AGTAAACAAGAGGGCCAGATTAACCAAAACTACTATCTTTGTACCAATTGCTTTT	27454
Qy	5671	GCTGAACAGGAGGGGGTGTGTGAGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT	5730
Db	27453	GCTGAACCAAGA-----	27442
Qy	5731	GT	5790
Db	27441	-----GTGAGTGAAGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	27392
Qy	5791	TGATATTTGGCCCTCATGCCACAGGGGTAATTGGGAGAGTC-----AGCTTTCTGC	5841
Db	27391	TGATATTTGGGGCCCTATGCCAAAGAGGGGTAATCAGTATGAGTCAAGGCTGTAAAGCTTCTGC	27332
Qy	5842	AAACACAGTAAGCTCCCAAGATGTGATTTGTGTGCTGTAATCCAAAGGGGAGGGCGATC	5901
Db	27331	AAACACAGCAGAGCTCCCAAGATTAATTTGTGTGCTGTAATCCAAAGGGGAGGGCGATC	27272
Qy	5902	AGAGTGCACAGAACTCAACAGATTAAGCCACCTGTGGGGCTCAGAGAGGAGGTTTACA	5961
Db	27271	AGAGTGCACAGAA-----CAACAAGATTAAGCCACCTGTGAGGGCTCAAG-----GAGTTTACA	27220
Qy	5962	AGAGGTTAAAGCCCAAGCATTATTTATTCACAGACATGACTCAAAATCAAAGTGCAAGAG	6021
Db	27219	AGAGGTTAAAGCCCAAGCATTATTTATTCACAGACATGACTCAAAATCAAAGTGCAAGAG	27160
Qy	6022	AGATTAGCTGAGAAATGGGGGTGTGACGTGTGGGACACTGTGACCTTGCACTTATATGTCA	6081
Db	27159	AGATTAGCTGAGAAATGGGGGTGTGACGTGTGGGATTAATCTGTGACTTGTGCCCTTTGATGTCA	27100
Qy	6082	CTAGGCCAAGAGAGCATGACACAGAGGATGACTGGATCTACTAGCTTTGAGCAGGACAGT	6141
Db	27099	CTAGGCCAAGAGAGCATGACACAGAGGATGACTGATCTACTAGCTTTGAGCAGGACAGT	27040
Qy	6142	GGAGAAATGGGTACTCTCATTCCTGATGTGAGAGGCTGAGACACACAGGTACAGTGTTC	6201
Db	27039	AGAGAAATGGG-----TGTGTGAGAGGTTTGTGAATAGACAGGATGCAAGGTTC	26992

Oy	6202 C-	-CTGTGTCATCATGCCAGATTCTTCCGCCACGTTTTCATAAGACTAAGCACTACTCTCT	6258
Db	26991	CTCGATGTGCCTCATTGCTGGAGTTCCAGCACGTTTTTCAAAGACCACGACTACTCTCN	26932
Oy	6259	GGTGAACAACAAAGTATCCAAGCCCTTAAGCCCATTTTGSTCTAATTAAATCAAGAACCCT	6318
Db	26931	NNN	26872
Oy	6319	GGGGATGCAGGCTCTGACAGCAGCAGGACTTTTAAAAAGCT	6359
Db	26871	NNNTT	26831
RESULT 6			
AC119514/c			
LOCUS			
DEFINITION	AC119514	80578 bp DNA linear HTG-18-JUL-2002	
ACCESSION	AC119514		
VERSION	AC119514.3	GI:21746577	
KEYWORDS	HTG; HTGS-PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 80578)		
AUTHORS	Munyir,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrook,S.L., Amarutunge,H.C., Are,J., Ayele,M., Banks,T., Barbati,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Burch,P., Burke,S., Brevet,M., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carrott,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,C., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davis,M.L., Davis,C., Davy-Carroll,L., Dedertich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,X., Dinh,H.H., Douthevalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Galisti,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,C., Kratovic,J., Kureshi,A., Landry,N., Leal-B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisgeed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Milner,G., Milner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S., Ogun,M., Okunode,G., Orangunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,K., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubkan,I., Rolfe,M., Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shooshatali,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swatek,A., Tabor,P., Tameris,A., Tamerlis,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanai,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 80578)		
AUTHORS	Worley,K.C.		
TITLE	Direct Submission		



Dh	59131	TACATCGAGGTCGAACACTGCACATATAAAATTAATAAATCAAAAATATATGT	59062
Oy	2081	GCGTTGGTAAACATTGTCTTGACATGTGATTGAAGCATGTACACGCCACACATGA	2140
Dh	59061	GGTATATGTGTGTGTATTAACACACACACTTATATATATATTTTTAGTCACAGCTTTCTGT	59002
Oy	2141	AGAG-----GGATGTGGGCGCGAGAGATGGCTCAGCGGTTA	2177
Dh	59001	GTAGCCCTGTGGTTTCTTAGAACACTACTCTCTAGGGGGCGGAGAGAGTGGCTCAGCGGTTA	58942
Oy	2178	AGAGCACTGACTGCTCTTCCGAGAAGAGTCTCGAGTTCATAATCTAGCACACCACATGGT	2237
Dh	58941	AGAGCACTGACTGCTCTTCC-----AGAGCTCTGAGTTCCAATCCACAGAACCATATGGT	58886
Oy	2238	GGCTCAACAACATCATATAGATGATCGACACCCCTCTGTGGATCGATCGAAGACAGCTG	2297
Dh	58885	GGCTCAACAACATCTGTATGTAGGATGTGTAGACCTCTCTGTGTGTGTGAAGAC-----	58831
Oy	2298	CAGAGCTACAGTACTAGATAGATATCTATAATAATATCTTTTTTAAAAAAA	2350
Dh	58830	---GGCTACAGCTACTATATATATATAATCAATAAATCTTTTTTAAAAAAA	58781
RESULT 7			
LOCUS	AC094169	181772 bp	DNA linear HTG 24-AUG-2002
DEFINITION	Rattus norvegicus clone CH230-2B3,	***	SEQUENCING IN PROGRESS ***
ACCESSION	AC094169		
VERSION	AC094169.4	GI:22094310	
KEYWORDS	HTG; HTGS; PHASEL.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 181772)		
AUTHORS	Muzny,D.,Kerle,,Metzker,M.,Lee,,Abramzon,S.,Adams,C.,Alder,J., Allen,C.,Allen,H.,Alsbrooks,S.,Amiri,A.,Angiano,D., Anyaldebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H., Baldwin,D.,Bandarainake,D.,Barber,M.,Banstead,M.,Benahmed,F., Biswalto,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M., Bryant,N.,Buhaý,C.,Burgh,J.,Burrell,K.,Calderon,E., Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A., Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J., Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L., Davila,M.L.,Davis,C.,Davy-Carrillo,L.,De Anda,C.,Dederich,D., Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K., Draper,H.,Dugan-Rocha,S.,Dunn,A.,Duplin,K.,Duval,B.,Eaves,K., Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G., Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P., Fraser,C.G.,Gabisi,A.,Ganta,R.,García,A., Garner,T.,Garza,M., Georgescu,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W., Gunaratne,P.,Haaland,M.,Hamill,C.,Hamilton,C., Hamilton,K., Harvey,Y.,Havlik,P.,Hawes,A.,Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huliy,S., Hume,J., Idolebird,D., Jackson,A., Jackson,L., Jacob,D., Jiang,H., Johnson,B.,Johnson,R.,Jolivet,A., Kapathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowals,C., Krefit,C.L., Lebow,H.,Levan,J.,Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louisseged,H., Losado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud M., Malloy,K., Mangum,A., Manungu,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathney,S., McLeod,M.,McNeill,T.,Meenen,E.,Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankevis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwoekelennu,O., Okwuonu,G., Olurunposoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Polidexter,A., Popovic,K., Primus,E., Pu,L.-U., Puzio,M., Quitozy,J., Rachlin,E., Reeves,K., Regier,M.A., Reighn.R., Reilly,B., Reilly,M., Ren,Y.,		

REFERENCE	Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajd, D., Sneed, A., Sodergren, E., Song, X., -Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmati, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Mu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, Y., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.
REFERENCE	Unpublished
REFERENCE	2 (bases 1 to 181772)
REFERENCE	Worley, K. C.
REFERENCE	Submitted (11-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 181772)
REFERENCE	Rat Genome Sequencing Consortium.
REFERENCE	Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	On Aug 3, 2002 this sequence version replaced gi:21700341.
REFERENCE	----- Genome Center
REFERENCE	Center: Baylor College of Medicine
REFERENCE	Center code: BCM
REFERENCE	Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>
REFERENCE	Contact: hgsc-help@bcm.tmc.edu
REFERENCE	----- Project Information
REFERENCE	Center project name: GACX
REFERENCE	Center clone name: CH230-2B3
REFERENCE	----- Summary Statistics
REFERENCE	Sequencing vector: Plasmid;
REFERENCE	Chemistry: Dye-terminator Big Dye; 100% of reads
REFERENCE	Assembly program: Phrap; version 0.990329
REFERENCE	Consensus quality: 135179 bases at least Q40
REFERENCE	Consensus quality: 140704 bases at least Q30
REFERENCE	Consensus quality: 145423 bases at least Q20
REFERENCE	-----
REFERENCE	* NOTE: Estimated insert size may differ from sequence length
REFERENCE	* (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> )
REFERENCE	* NOTE: This is a 'working draft' sequence. It currently
REFERENCE	* consists of 55 contigs. The true order of the pieces
REFERENCE	* is not known and their order in this sequence record is
REFERENCE	* arbitrary. Gaps between the contigs are represented as
REFERENCE	* runs of N, but the exact sizes of the gaps are unknown.
REFERENCE	* This record will be updated with the finished sequence
REFERENCE	* as soon as it is available and the accession number will
REFERENCE	* be preserved.
REFERENCE	1 1210: contig of 1210 bp in length
REFERENCE	1211 1310: gap of unknown length
REFERENCE	1311 2908: contig of 1598 bp in length
REFERENCE	2909 3008: gap of unknown length
REFERENCE	3009 4605: contig of 1597 bp in length
REFERENCE	4606 4705: gap of unknown length
REFERENCE	4706 6407: contig of 1702 bp in length
REFERENCE	6408 6507: gap of unknown length
REFERENCE	6508 7760: contig of 1253 bp in length
REFERENCE	7761 7860: gap of unknown length
REFERENCE	7861 9288: contig of 1428 bp in length
REFERENCE	9289 9388: gap of unknown length
REFERENCE	9389 10415: contig of 1027 bp in length
REFERENCE	10416 10515: gap of unknown length
REFERENCE	10516 12123: contig of 1608 bp in length
REFERENCE	12124 12223: gap of unknown length
REFERENCE	12224 13352: contig of 1129 bp in length
REFERENCE	13353 13452: gap of unknown length
REFERENCE	13453 15155: contig of 1703 bp in length

*	15156	15255:	gap of unknown length
*	15256	16618:	contig of 1363 bp in length
*	15619	16718:	gap of unknown length
*	16719	18622:	contig of 1904 bp in length
*	18623	18722:	gap of unknown length
*	18723	19829:	contig of 1107 bp in length
*	19830	19929:	gap of unknown length
*	19930	21969:	contig of 2040 bp in length
*	21970	22069:	gap of unknown length
*	22070	22915:	contig of 1846 bp in length
*	22916	24015:	gap of unknown length
*	24016	26059:	contig of 2044 bp in length
*	26060	26159:	gap of unknown length
*	26160	28222:	contig of 2063 bp in length
*	28223	30399:	contig of 2077 bp in length
*	30400	30499:	gap of unknown length
*	30500	32655:	contig of 2156 bp in length
*	32656	32755:	gap of unknown length
*	32756	33854:	contig of 1099 bp in length
*	33855	33954:	gap of unknown length
*	33955	35237:	contig of 1283 bp in length
*	35238	35337:	gap of unknown length
*	35338	36771:	contig of 1434 bp in length
*	36772	36871:	gap of unknown length
*	36872	38777:	contig of 1906 bp in length
*	38778	38877:	gap of unknown length
*	38878	40669:	contig of 1792 bp in length
*	40670	40769:	gap of unknown length
*	40770	43169:	contig of 2400 bp in length
*	43170	43269:	gap of unknown length
*	43370	45568:	contig of 2299 bp in length
*	45569	45668:	gap of unknown length
*	45669	47568:	contig of 2190 bp in length
*	47569	47959:	gap of unknown length
*	47959	49767:	contig of 1809 bp in length
*	49768	49867:	gap of unknown length
*	49868	52667:	contig of 2700 bp in length
*	52668	52667:	gap of unknown length
*	52669	55159:	contig of 2492 bp in length
*	55160	55259:	gap of unknown length
*	55260	57887:	contig of 2628 bp in length
*	57888	57987:	gap of unknown length
*	57988	60041:	contig of 2054 bp in length
*	60042	60141:	gap of unknown length
*	60142	62866:	contig of 2725 bp in length
*	62867	62966:	gap of unknown length
*	62967	66014:	contig of 3048 bp in length
*	66015	66114:	gap of unknown length
*	66115	68172:	contig of 2058 bp in length
*	68173	68272:	gap of unknown length
*	68273	72352:	contig of 4080 bp in length
*	72353	72452:	gap of unknown length
*	72453	76168:	contig of 3716 bp in length
*	76169	76268:	gap of unknown length
*	76269	79795:	contig of 3521 bp in length
*	79796	79895:	gap of unknown length
*	79896	84438:	contig of 4543 bp in length
*	84439	84538:	gap of unknown length
*	84539	88636:	contig of 4118 bp in length
*	88637	88756:	gap of unknown length
*	88757	93275:	contig of 4519 bp in length
*	93276	93376:	gap of unknown length
*	93377	98043:	contig of 4668 bp in length
*	98044	98143:	gap of unknown length
*	98144	103024:	contig of 4881 bp in length
*	103025	103124:	gap of unknown length
*	103125	107830:	contig of 4706 bp in length
*	107831	107930:	gap of unknown length
*	107931	114501:	contig of 6571 bp in length
*	114502	114601:	gap of unknown length
*	114602	119361:	contig of 4760 bp in length
*	119362	119361:	gap of unknown length

[illegible]

Barbata, J., Benton, J., Blame, K., Blankenburg, K., Bonaldi, D.,  
 Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Eathat, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunatirane, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,  
 Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Krstovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunonu, G.,  
 Orangun, N., Oyiedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojuben, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,  
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Syatek, A., Taber, P., Tameris, A., Tameris, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Woodem, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G. and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 194630)  
 Morley, K.C.

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
 Center project name: GYO  
 Center clone name: CH230-152K9

----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 135641 bases at least Q40  
 Consensus quality: 141731 bases at least Q30  
 Consensus quality: 146942 bases at least Q20

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_difft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_difft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 59 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1793: contig of 1793 bp in length

1794 1893: gap of unknown length  
 1894 3132: contig of 1239 bp in length  
 3133 3232: gap of unknown length  
 3233 4328: contig of 1096 bp in length  
 4329 4428: gap of unknown length  
 4429 5580: contig of 1152 bp in length  
 5581 5681: gap of unknown length  
 5681 7470: contig of 1790 bp in length  
 7471 7570: gap of unknown length  
 7571 9126: contig of 1556 bp in length  
 9127 9226: gap of unknown length  
 9227 10595: contig of 1369 bp in length  
 10596 10695: gap of unknown length  
 10696 11883: contig of 1188 bp in length  
 11884 11983: gap of unknown length  
 11984 13414: contig of 1431 bp in length  
 13415 14925: contig of 1411 bp in length  
 14926 15025: gap of unknown length  
 15026 16379: contig of 1354 bp in length  
 16380 16479: gap of unknown length  
 16480 17657: contig of 1178 bp in length  
 17658 17757: gap of unknown length  
 17758 18831: contig of 1074 bp in length  
 18832 18931: gap of unknown length  
 18932 20123: contig of 1192 bp in length  
 20124 20223: gap of unknown length  
 20224 21898: contig of 1675 bp in length  
 21899 23422: gap of unknown length  
 23423 23523: gap of unknown length  
 23523 25401: contig of 1879 bp in length  
 25402 25501: gap of unknown length  
 25502 27167: contig of 1666 bp in length  
 27168 28730: contig of 1463 bp in length  
 28731 28830: gap of unknown length  
 28831 30823: contig of 1993 bp in length  
 30824 30923: gap of unknown length  
 30924 32988: contig of 2065 bp in length  
 32989 33088: gap of unknown length  
 33089 35579: contig of 2491 bp in length  
 35580 35679: gap of unknown length  
 35680 37193: contig of 1514 bp in length  
 37194 37293: gap of unknown length  
 37294 38574: contig of 1281 bp in length  
 38575 38674: gap of unknown length  
 38675 40473: contig of 1799 bp in length  
 40474 40573: gap of unknown length  
 40574 42832: contig of 2259 bp in length  
 42833 42932: gap of unknown length  
 42933 44836: contig of 1904 bp in length  
 44837 44936: gap of unknown length  
 44937 46922: contig of 1866 bp in length  
 46923 47022: gap of unknown length  
 47023 48984: contig of 1662 bp in length  
 48985 49084: gap of unknown length  
 49085 52798: contig of 3714 bp in length  
 52799 52898: gap of unknown length  
 52899 55863: contig of 2965 bp in length  
 55864 55963: gap of unknown length  
 55964 57728: contig of 1765 bp in length  
 57729 57828: gap of unknown length  
 57829 59470: contig of 1642 bp in length  
 59471 59570: gap of unknown length  
 59571 62556: contig of 2986 bp in length  
 62557 62657: gap of unknown length  
 62657 65039: contig of 2283 bp in length  
 65040 65139: gap of unknown length  
 65140 67858: contig of 2719 bp in length  
 67859 70147: gap of unknown length  
 70148 70247: contig of 2189 bp in length

*	70248	72929:	contig of 2682 bp in length
*	72930	73029:	gap of unknown length
*	73030	76076:	contig of 3047 bp in length
*	76077	71676:	gap of unknown length
*	76177	79311:	contig of 3135 bp in length
*	79312	79411:	gap of unknown length
*	79412	82900:	contig of 3489 bp in length
*	82901	83000:	gap of unknown length
*	83001	87434:	contig of 4434 bp in length
*	87435	87534:	gap of unknown length
*	87535	92930:	contig of 5396 bp in length
*	92931	93030:	gap of unknown length
*	93031	96209:	contig of 3179 bp in length
*	96210	96309:	gap of unknown length
*	96310	99070:	contig of 2761 bp in length
*	99071	99170:	gap of unknown length
*	99171	102188:	contig of 3018 bp in length
*	102189	102288:	gap of unknown length
*	102289	105629:	contig of 3341 bp in length
*	105630	105729:	gap of unknown length
*	105730	111114:	contig of 5385 bp in length
*	111115	111214:	gap of unknown length
*	111215	114351:	contig of 3137 bp in length
*	114352	114451:	gap of unknown length
*	114452	119170:	contig of 4719 bp in length
*	119171	119270:	gap of unknown length
*	119271	123696:	contig of 4426 bp in length
*	123697	123796:	gap of unknown length
*	123797	133294:	contig of 9498 bp in length
*	133295	133394:	gap of unknown length
*	133395	141172:	contig of 7778 bp in length
*	141173	141272:	gap of unknown length
*	141273	148661:	contig of 7389 bp in length
*	148662	148761:	gap of unknown length
*	148762	157010:	contig of 8249 bp in length
*	157011	157110:	gap of unknown length
*	157111	166469:	contig of 9359 bp in length
*	166470	166569:	gap of unknown length
*	166570	175385:	contig of 8816 bp in length
*	175386	175485:	gap of unknown length

Query Match	2.1%	Score 188.6	DB 2	Length 194630
Best Local Similarity	67.4%	Pred. No. 2.4e+43		
Matches 359	Conservative 0	Mismatches 134	Indels 40	Gaps 5
QY 1846	ATATGTTGGGAGGAGAGATGGCTCAGGCTCCAGAGGACCTTGCTCTTCAGAGAGAC	1905		
Db 74564	ATTGGGGGGGGCTGGAGATGGTTCAGTGTTAGGATGCTACGCTGGCTCTTCACAGAGAT	74505		
QY 1966	CTAGTTTACGTTCCAGGACATCATATGGTGGTCCACAGGACATCTGTAATCCAGTTCGAG	1965		
Db 74504	GTGGGCTACCTTTTCAGTACCCATATGAGGGCTCCACAACTTCTGTAATCCAGTTCGAT	74445		
QY 1966	AGGGTTCACACACCTCTTCTGGCTCCACAGGAC-----CACATCATATGATACACAGACA	2021		
Db 74444	GGGATCCAGTACCATCTCTGACTTCCCTGGGGCCCGAGGACATGATACACAGACA	74385		
QY 2022	TACATGACAGGCAAAACCCCATACACACATAA-ATTAATTAAGAAACTTAAAGGTGCAT	2080		
Db 74384	TACATGACAGGCTCAAAACCTCATGACATATAAATAAATCAATAATCAAAAATATATGT	74325		
QY 2081	GTCGTTGGAACATGTGTCCTTACACATGCTGATTTGAAGCATGTACAACGACACACATGA	2140		
Db 74324	GTGTAATGTGTGTATTTACACACACACTTATATATTTATATTTTTTACGACGCTTTTCTGT	74265		
QY 2141	AGAG-----GGATCTGGGGCTGGAGAGATGGCTCAGCGGTTA	2177		
Db 74264	GTAACCCCTGGTTTTCTTGAAGACTCACTCTCTAGGGGGGCTGGAGAGGTGGCTCAGCGGTTA	74205		
QY 2178	AGAGCAGTCACTGCTCTTCCAGGAAGGCTCTGAGTTCAAAATCTAGCAACCATGTGT	2237		
Db 74204	AGAGCAGTCACTGCTCTTCC-----AGAGCTCTGAGTTCAAAATCCAGCAACCATGTGT	74149		

Oy	2238	GGCTACACAACCATTCATTAATGAGATCTACACACCCCTTCTTGCGATCTGAAGACAGCTG	2297
Db	74148	GGCTCACCAACCATCTGTATATGGGAACTATGACCCCTTCTGGTGTTGAAGAC----	74094
Oy	2298	CAGAGCTCAGCTGACTAGATAATFACATAAATAAATCTTTTTTAAAAAAA	2350
Db	74093	---GGCTCAGCTCCTACTTATATATATAATCAATCAATAATCTTTTTAAAAAAA	74044
RESULT 9			
ACOL13320			
LOCUS		149425 bp DNA linear HTG 18-NOV-2000	
DEFINITION		Homo sapiens chromosome 2 clone RP11-36003 map 2, WORKING DRAFT	
ACCESSION		ACOL13320	
VERSION		ACOL13320.7 GI:11225433	
KEYWORDS		HTG; HTGS; PHASEI; HTGS_DRAFT.	
SOURCE		Homo sapiens.	
ORGANISM		Homo sapiens.	
		Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE		1 (bases 1 to 149425)	
AUTHORS		Birten,B., Linton,L., Nusbaum,C. and Lander,E.	
TITLE		Homo sapiens chromosome 2, clone RP11-36003	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 149425)	
AUTHORS		Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,	

TITLE  
JOURNAL  
COMMENT

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

----- Genome Center  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

```
***** Project Information
Center project name: L3776
```

```
Center clone name: 360_O_3
----- Summary Statistics
```

Sequencing vector: M13; M77815; 12% of reads  
Sequencing vector: Plasmid; n/a; 88% of reads

Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap: version 0.960731

Consensus quality: 140887 bases at least Q40  
Consensus quality: 144879 bases at least Q30

Consensus quality: 146417 bases at least Q20  
Insert size: 328000: agarose-fn

Insert size: 22000; agarose 1p  
Insert size: 147625; sum-of-contigs  
Overall coverage: 3.6 in 920 bases; average-fn

Quality coverage: 5.5 in Q20 ba.

\* NOTE: THIS IS A WORKING DRAFT SEQUENCE. IT CURRENTLY CONSISTS OF 19 CONTIGS. THE TRUE ORDER OF THE PIECES

- \* Is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

- \* runs of N, but the exact sizes of the gaps are unknown
- \* This record will be updated with the finished sequence





Db 45479 GCCTCTAGCATATGACCAACACCACGACA 45510

RESULT 10  
AC021016 191754 bp DNA linear PRI 09-MAY-2001  
LOCUS AC021016 Homo sapiens BAC clone RP11-378A13 from 2, complete sequence.  
DEFINITION AC021016 AC021016 GI:11120952  
VERSION AC021016.4  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 191754)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 191754)  
AUTHORS Cordes, M., Maupin, R., Hawkins, M. and Boyer, E.  
TITLE The sequence of Homo sapiens BAC clone RP11-378A13  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 191754)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (12-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 191754)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 5 (bases 1 to 191754)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 8, 2000 this sequence version replaced gi:7650969.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0378A13  
-----  
NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.  
  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.  
  
MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>  
  
SOURCE INFORMATION:  
The RPc1-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frenken, E.,  
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved

approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at the Roswell Park Cancer Institute  
(<http://bacpac.med.buffalo.edu>)  
VECTOR: pBACe3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP111-146N10; the clone  
sequenced to the right is RP11-36003. Actual start of this clone  
is at base position 1 of RP11-378A13; actual end is at base  
position 191754 of RP11-378A13.

The sequence RP11-378A13 from 156426 to 156433 is derived from a  
single ml3 subclone. Size of the region was confirmed by PCR from  
BAC DNA.

FEATURES  
source location/Qualifiers  
1..191754  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-378A13"  
/clone\_lib="RPc1-11"  
1..224  
/note="similar to EST AA947005 (NID:g3110400) oq45e10.s1"  
1..224  
/note="similar to EST A1184556 (NID:g3735194) qd60a10.x1"  
1..224  
/note="similar to EST A157236 (NID:g4310105) tk01d06.x1"  
1..216  
/note="similar to EST AA782751 (NID:g2842082) aj08c02.s1"  
1..216  
/note="similar to EST AW020425 (NID:g5873955) df09e11.y1"  
2..223  
/note="similar to EST AA175052 (NID:g1756173) ms82f02.r1"  
2..223  
/note="similar to EST AV079022 (NID:g5198850) "  
2..223  
/note="similar to EST AW538472 (NID:g7180889) "  
2..49  
/note="similar to EST AV330504 (NID:g6370956) "  
7..224  
/note="similar to EST A154294 (NID:g4292719) "  
7..224  
/note="similar to EST AW433991 (NID:g9656298) "  
12..223  
/note="similar to EST AW059787 (NID:g6652109) "  
20..223  
/note="similar to EST A1709077 (NID:g4998853) as86d11.x1"  
60..220  
/note="similar to EST AA870364 (NID:g2965809) vq44e05.r1"  
67..223  
/note="similar to EST AV330504 (NID:g6370956) "  
447..660  
/rpt\_family="Alu"  
1182..1425  
/rpt\_family="L1"  
1442..1583  
/rpt\_family="MERL-type"  
1820..1928  
/rpt\_family="Alu"  
1948..2347  
/note="similar to EST AA224009 (NID:g1844621) zr12g06.r1"  
2162..2293  
/rpt\_family="Alu"  
2297..2587  
/rpt\_family="Alu"  
2597..2804  
/rpt\_family="Alu"  
2606..2923  
/note="similar to EST AW351985 (NID:g6849698) "  
2630..3135  
/note="similar to EST AA234699 (NID:g1859192) zs39g05.r1"

	misc_feature	2804..3242	/note="similar to	EST AI916611 (NID:g5636466)	wz28f08.x1"
	misc_feature	2976..3245	/note="similar to	EST AI401669 (NID:g4244756)	lh24e10.x1"
	repeat_region	3097..3284	/rpl_family="MIR"		
	misc_feature	3546..3781	/note="similar to	EST AA102341 (NID:g1647297)	zm26h05.r1"
	repeat_region	3848..4127	/rpl_family="Alu"		
	repeat_region	4187..4469	/rpl_family="Alu"		
	repeat_region	4474..4690	/rpl_family="Alu"		
	misc_feature	4723..5241	/note="similar to	EST AM444974 (NID:g6986736)"	
	misc_feature	4931..5413	/note="similar to	EST AA100813 (NID:g1647248)	zm26h05.s1"
	repeat_region	6006..6137	/rpl_family="MIR"		
	repeat_region	6243..6394	/rpl_family="MIR"		
	misc_feature	6967..7092	/note="similar to	EST AW461516 (NID:g7031684)"	
	misc_feature	7935..8054	/note="similar to	EST AW461516 (NID:g7031684)"	
	repeat_region	8294..8433	/rpl_family="MERL_type"		
	misc_feature	8553..8604	/note="similar to	EST AI391439 (NID:g4217443)	tf96e02.x1"
	misc_feature	8556..8749	/note="similar to	EST AW461516 (NID:g7031684)"	
	misc_feature	9362..9733	/note="similar to	EST AI391439 (NID:g4217443)	tf96e02.x1"
	misc_feature	9476..9723	/note="similar to	EST AA595679 (NID:g2411029)	nk66f01.s1"
	misc_feature	9997..10265	/note="similar to	EST AI634626 (NID:g4685956)	tz30g01.x1"
	misc_feature	10003..10482	/note="similar to	EST AI828100 (NID:g5448771)	wk31g06.x1"
	misc_feature	10003..10482	/note="similar to	EST AI890574 (NID:g5585738)	wm87e04.x1"
	misc_feature	10003..10482	/note="similar to	EST AI978960 (NID:g5803990)	wr68e06.x1"
	misc_feature	10003..10482	/note="similar to	EST AL047914 (NID:g4728747)"	
	misc_feature	10003..10482	/note="similar to	EST AW027349 (NID:g5886105)	wc73e09.x1"
	misc_feature	10003..10457	/note="similar to	EST AI972200 (NID:g5769026)	wr63h01.x1"
	misc_feature	10003..10421	/note="similar to	EST W48859 (NID:g1336988)	zc42c08.s1"
	misc_feature	10003..10330	/note="similar to	EST AW074215 (NID:g6029213)	xb09d12.x1"
	misc_feature	10003..10249	/note="similar to	EST AA143563 (NID:g1712935)	zo3zd10.s1"
	misc_feature	10003..10247	/note="similar to	EST AI583619 (NID:g4569516)	ts16d09.x1"
	misc_feature	10003..10240	/note="similar to		
Query Match		2.1%; Score 188; DB 9; Length 191754;			
Best Local Similarity		67.4%; Pred.No.3.6e-43;			
Matches 507; Conservative		0; Mismatches 190; Indels 55; Gaps 15;			
OY	3238	AGAGCCAAAGGGCGGCGACACGGCCTCACGCTTCAGGCTGTGCTGCGGCTCTAGGTT	3297		
	Db	164808	AAAAGCAAGGGTGGGGGCACA--TCAGCTCTCATTTGTCAAGGCG-TGGAGGCGCTTGCGCT	164864	
OY	3298	CCCAGGAGACTTGG--GCACCTTATCCCCACCACCCCACTTCATTCTCTGCGGGGCCATT	3355		
	Db	164865	CCGAGGACCTTGGCAGCAGCACTTCCCTCCCAACCCCTTCAC-----CTCTCTGGGGCCCTATT	164919	
OY	3356	CTTCCCTTATATGTGAAGAATTCTGTGGGGGGGGGGGTGTGTGTAAGCAAAAGTCT	3415		

Db 164920	CTTCCCTATATAGTGAAGGACGTTCTTAGAGAGAGGGG-----GGTGGGACAAAG--	164972
Qy 3416	GTTCGGTCTCCGACAGCCAGCTTGGCACAACTCTTAAGATCTCCAGAGTGGTGGCTGC	3475
Db 164973	--TGCCTCTTCTGACAGCCAGCTTGGCACAACTCTTCCAGATCTCCAGTGGACACTGCC	165030
Qy 3476	TCTTCCAGACAGGTAAAGCAATTTGGGTGGGACACATGGTGAACACAGTGGTGGAGG	3535
Db 165031	TCGCCAAGACAGGTAAAGCAACCTGGTGGAGACATGAGACCTTAGTGGTGGAGAGG	165090
Qy 3536	GACAGGGTCTT--TGCTTCTCTCTGGCAGCTT--GTGCTTTCTAGCAGCTTGGTATAGT	3593
Db 165091	GACGGGGTCTCTGCTTTTCTGGGCAACTTGGGGCTCTCTGAGACACTTGTGTACGGGC	165150
Qy 3594	TTGGGGGTGAGGTAAAGTGTCTCT--GAAACTCTGAAAGAACAAAGAGCC--AGCAGGCTG	3650
Db 165151	TGCAAGGGGAGGCTTGACACCTGTGGATACCTGAGAGAGTCTGGAGGCCAATGTGTCTG	165210
Qy 3651	TCTTGGGCTTCAATGAAGAAATTACAGAGCCCTCTCTCTGA--GTACACTTTCGT	3708
Db 165211	CCCTGGGCTTCCAGGGAAGACTTTTTCAGAGGCTTGGCCCAACATAGTCTCTAGTC	165270
Qy 3709	TGATCTGTGTAGATTTCCCTGGACCAAGTGTCT-----CTTGGACTAGATT	3757
Db 165271	TCTTCACTGTGTGCTAGCCCACTCCACTTTTGGGTGTACACATCCCAATTTGCCCTGGACT	165330
Qy 3758	TCTTCAATTAATTAAGTGAAGGACAGTCCCTGAG-----ACTTGGATCTGGTCCCTGT	3805
Db 165331	TCTTGAAGTTAAACAAGACACATGTCTGGGACAGACCGGAGCTGTGGTGTCTCTTAACCTC	165390
Qy 3806	ATTTAC---TACTTCTCTGTGCTCTCACTTCTGTTCATGATCTTACATCTGAAT	3862
Db 165391	ACTCCCGACGCCCCCTCCACACATCTCTTTTGGCCGTGTCTTATCTGAAGT	165450
Qy 3863	GG---TTTCTTTGTGTACCAATTCCTCTGACACTCCTGGAGGTGTATCTTGGACAT	3919
Db 165451	GTTTTTTTCTTTGTGTACCAATTCCTCTGACACTCCTGGAGCTCCGGTCTCTGCTCTCT	165510
Qy 3920	GTATCTCTGGATGTAAAGCTGCGACGACCAAGCA 3951	
Db 165511	GGCTCTTACATATGAGACAGCAACCAAGCA 165542	
RESULT 11	AC115965	220999 bp DNA linear HTG 17-JUN-2002
LOCUS	Mus musculus clone RP24-69114, WORKING DRAFT SEQUENCE, 8 ordered	
DEFINITION	pieces.	
ACCESSION	AC115965	
VERSION	AC115965.2 GI:21431261	
KEYWORDS	HTG: HTGS_PHA5E2; HTGS_DRAFT; HTGS_FULLTOP.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Mumukhya, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	1 (bases 1 to 220909)	
TITLE	Birren, B., Linton, L., Nushum, C. and Lander, E.	
JOURNAL	Mus musculus, clone RP24-69114	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 220909)	
REFERENCE	Birren, B., Linton, L., Nushum, C., Lander, E., All, A., Allen, N.,	
AUTHORS	Anderson, S., Barra, N., Bastien, V., Bloom, T., Boguslavsky, L.,	
REFERENCE	Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,	
AUTHORS	Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,	
REFERENCE	Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,	
AUTHORS	Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,	
REFERENCE	Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,	
AUTHORS	Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,	
REFERENCE	Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R.,	
AUTHORS	Landers, T., Lehoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,	
REFERENCE	Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,	
AUTHORS	McCarthy, M., McEwan, P., McKernan, K., Melidim, J., Menus, L.,	



Db 119244 TGAATCTGATCCCTCTCTGAGTGTCTGAAGAC-----AGCTACGTACTCTTA 119295  
 QY 2317 GATATCTACTATATAATATCTTTTAAAAAAA 2350  
 Db 119296 CATATATAATATAATATCTTTAAAAAAA 119329

RESULT 12  
 AC102777/c  
 LOCUS AC102777 184591 bp DNA linear HTG 21-AUG-2002  
 DEFINITION Mus musculus clone RP23-5K7, WORKING DRAFT SEQUENCE, 25 unordered  
 pieces.  
 AC102777  
 AC102777.2 GI:22381791  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 184591)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP23-5K7  
 Unpublished  
 2 (bases 1 to 184591)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,  
 Anderson,S., Barra,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Kells,C., Iarocque,K.,  
 Lamazates,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarty,M., McEwan,P., McKernan,K., McSheeters,R., Meldrum,J.,  
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phukhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Roselt,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 184591)  
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,  
 Barra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarty,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced gi:17060887.  
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L19649  
 Center clone name: 5\_K-7  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 176248 bases at least Q40  
 Consensus quality: 180182 bases at least Q30  
 Consensus quality: 181502 bases at least Q20  
 Insert size: 188000; agarose-fp  
 Insert size: 182191; sum-of-coverage  
 Quality coverage: 5.8 in Q20 bases; agarose-fp  
 Quality coverage: 6.0 in Q20 bases; sum-of-coverage  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 25 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 806: contig of 806 bp in length  
 \* 807 906: gap of 100 bp  
 \* 907 1108: contig of 202 bp in length  
 \* 1109 1208: gap of 100 bp  
 \* 1209 1890: contig of 682 bp in length  
 \* 1891 1990: gap of 100 bp  
 \* 1991 2684: contig of 694 bp in length  
 \* 2685 2784: gap of 100 bp  
 \* 2785 3648: contig of 864 bp in length  
 \* 3649 3748: gap of 100 bp  
 \* 3749 5130: contig of 1382 bp in length  
 \* 5131 5230: gap of 100 bp  
 \* 5231 6203: contig of 973 bp in length  
 \* 6204 6303: gap of 100 bp  
 \* 6304 7763: contig of 1460 bp in length  
 \* 7764 7863: gap of 100 bp  
 \* 7864 8416: contig of 553 bp in length  
 \* 8417 8516: gap of 100 bp  
 \* 8517 11254: contig of 2738 bp in length  
 \* 11255 11354: gap of 100 bp  
 \* 11355 123417: contig of 2063 bp in length  
 \* 13418 13517: gap of 100 bp  
 \* 13518 15594: contig of 2077 bp in length  
 \* 15595 15694: gap of 100 bp  
 \* 15695 19567: contig of 3873 bp in length  
 \* 19568 19667: gap of 100 bp  
 \* 19668 22294: contig of 2627 bp in length  
 \* 22295 22394: gap of 100 bp  
 \* 22395 26269: contig of 3875 bp in length  
 \* 26270 26369: gap of 100 bp  
 \* 26370 29166: contig of 2797 bp in length  
 \* 29167 29266: gap of 100 bp  
 \* 29267 34448: contig of 5182 bp in length  
 \* 34449 34548: gap of 100 bp  
 \* 34549 43348: contig of 8800 bp in length  
 \* 43349 43448: gap of 100 bp  
 \* 43449 51979: contig of 8531 bp in length  
 \* 51980 52079: gap of 100 bp  
 \* 52080 67532: contig of 15453 bp in length  
 \* 67533 67632: gap of 100 bp  
 \* 67633 79233: contig of 11601 bp in length  
 \* 79234 79333: gap of 100 bp  
 \* 79334 96823: contig of 17490 bp in length

*	96924	96923: gap of	100 bp
*	96924	122005: contig of	25082 bp in length
*	122006	122105: gap of	100 bp
*	121206	161850: contig of	39745 bp in length
*	161851	161950: gap of	100 bp
*	161951	184551: contig of	22641 bp in length

FEATURES	Location/Qualifiers
source	1. .184591

misc_feature	/olone.lib="RPCT-23 Female Mouse BAC"
1..806	
misc_feature	/note="assembly_fragment"
907..1108	
misc_feature	/note="assembly_fragment"
1209..1890	
misc_feature	/note="assembly_fragment"
1391..2684	
misc_feature	/note="assembly_fragment"
2785..3648	
misc_feature	/note="assembly_fragment"
3749..5130	
misc_feature	/note="assembly_fragment"
5231..6203	
misc_feature	/note="assembly_fragment"
6304..7763	
misc_feature	/note="assembly_fragment"
7864..8416	
misc_feature	/note="assembly_fragment"
8517..11254	
misc_feature	/note="assembly_fragment"
11355..13417	
misc_feature	/note="assembly_fragment"
13518..15594	
misc_feature	/note="assembly_fragment"
15695..19567	
misc_feature	/note="assembly_fragment"
19668..22294	
misc_feature	/note="assembly_fragment"
22395..26269	
misc_feature	/note="assembly_fragment"
26370..29166	
misc_feature	/note="assembly_fragment"
29267..34448	
misc_feature	/note="assembly_fragment"
34549..43348	
misc_feature	/note="assembly_fragment"
43449..51979	
misc_feature	/note="assembly_fragment"
52080..67532	
misc_feature	/note="assembly_fragment"
67633..79233	
misc_feature	/note="assembly_fragment"
79334..96823	
misc_feature	/note="assembly_fragment"
96924..122005	
misc_feature	/note="assembly_fragment"
122106..161850	
misc_feature	/note="assembly_fragment"
161951..184591	
misc_feature	/note="assembly_fragment"
clone_end:17	

Query Match	2.0%;	Score 184;	DB 2;	Length 184591;
Best Local Similarity	69.5%;	Pred. No. 5.3e-42;		
Matches 346;	Conservative 0;	Mismatches 125;	Indels 27;	Gaps 6

QY	1847	TATGTTGGGAGGAGATGGCTCCAGGCTTCCAGGAGGACACTGTCGTCCTTGGACAGAGAC	1906
Db	89227	TGTGATTCCTGTGCTTTAAGAGCTCACTGCTCAGAGGACACTGCTGCTCTTTCAGAGGACC	89166
QY	1907	TAGATTCAGTTCCAGGAGCTCATATGTTGGGCTCACAGGACATCTGTAATACGAGTTCCAGA	1966

[illegible]

DB	88773	AATAAATCTTTT	TTTTT	88756
Y	2320	ATATATATATATAT	ATATAT	2343

LOCUS	DEFINITION	AF100956/c	RESULT 13
AF100956	273800 bp DNA	linear	ROD 03-NOV-1998
	Mus musculus major histocompatibility locus class II region; Fas-binding protein Daxx (DAXX) gene, partial cds; Blnl1 (BlnG1), tapasin (tapasin), RalGDS-like factor (RLF), KE2 (KE2), BlnG4 (BlnG4), beta1, 3-galactosyl transferase (beta1,3-galactosyl transferase), ribosomal protein subunit S18 (RPS18), Sacm21 (Sacm21), H2-M1 (p) (H2-M1(p)), RING1 (RING1), KE6a (KE6a), KE4 (KE4), RXRbeta (RXRbeta), collagen alpha-2 (X1) (COL1A2), H2-O alpha (H2-Oalpha), RING3 (RING3), H2-M alpha (H2-M alpha), H2-M beta 2 (H2-M beta2), and H2-M beta1 (H2-M beta1) genes, complete cds; and LMP 2 gene, partial cds.		

ACCESSION	AF100956
VERSION	AF100956.1
	GI:3811374

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 273800)  
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**AUTHORS**  
Rowen, L., Qin, S., Madan, A., Loretz, C., James, R., Dots, M., Mix, L.,  
Hall, J., Lasky, S. and Hood, L.

TITLE Sequence of the mouse major histocompatibility locus class II region

JOURNAL REFERENCE	Unpublished 2 (bases 1 to 273800)
----------------------	--------------------------------------

AUTHORS	Rowen, L.
TITLE	Direct Submission

**JOURNAL**  
Submitted (26-OCT-1998) Department of Molecular Biotechnology, Box  
357730 University of Washington, Seattle, WA 98195, USA

**COMMENT** This sequence overlaps the entry in GenBank Accession Number AF027865 by 22022 bases. Sequencing methodology: high redundancy

Shotgun, using M3 templates. Interspersed Repeats were identified with RepeatMasker (available from

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

```
source 1. .273800
/organism="Mus musculus"
```

```
/strain="129svj"  
/db_xref="taxon:10090"
```

```

source
    /chromosome="17"
    1. .139758
    /organism="Mus musculus"
    /strain="129svj"
    /db_xref="taxon:10090"
    /clone="BAC 445D10"
    /clone.lib="Genome Systems"
    139672. .140387
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="BAC 346M2"
    /clone.lib="Research Genetics"
    /note="A PCR product made from BAC 346M2 was sequenced to
    fill the gap between BACs 445D10 and 36B7"
    140323. .273800
    /organism="Mus musculus"
    /strain="129svj"
    /db_xref="taxon:10090"
    /clone="BAC 36B7"
    /clone.lib="Genome Systems"
    <1. .2000
    /gene="DAXX"
    join(<146. .208,325. .544,1825. .2000)
    /product="Fas-binding protein Daxx"
    join(<146. .208,325. .544,1825. .1884)
    /gene="DAXX"
    /note="Intron-exon boundaries defined in relation to cDNA
    sequence in GenBank Accession Number AF006040"
    /codon_start=2
    /product="Fas-binding protein Daxx"
    /protein_id="AAC69891.1"
    /db_xref="GI:3811375"
    /translation="PPSKRFRKKQLQGLGNSYIREPMAGDSGNTSVQPMPS
    PLASVAVSDSTRVDSPELSHELVTSLCSPSPSLLLQTPQASLRQCIYKTSVATQCD
    PEELIVLSDSD"
    724. .858
    /rpt_family="MMB1"
    complement(1257. .1297)
    /rpt_family="(CA)n"
    1303. .1421
    /rpt_family="MMB1F"
    complement(1426. .1630)
    /rpt_family="B3"
    complement(1646. .1792)
    /rpt_family="MMB1F"
    complement(2728. .2788)
    /rpt_family="GC-rich"
    3168. .3214
    /rpt_family="MIR"
    <3365. .>5281
    /gene="BIN1"
    /note="zn finger; mouse orthologue to BIN1 l, described in
    GenBank Accession Number z97183"
    <3365. .>5281
    /product="Bin1"
    /gene="BIN1"
    3365. .5281
    /note="this orf was predicted by Genscan and confirmed by
    ESTs AA538332 and AA797063"
    /codon_start=1
    /product="Bin1"
    /protein_id="AAC69892.1"
    /db_xref="GI:3811376"
    /translation="MEPSALSPSGATLPLPLAPPLPLPAAVVHVSPFEVTSAL
    ESLNOORLOGLDQVSIROGREFRAHRAVLAASPPYHDVLLKGMTSISLPSMDP
    GAFETVLASATYGRISMAADIVNLTIVGSLVGMHIDKTELLREBSAATTTTV
    TAAPSVSPCASVPSGNGTVAPTVGSVSHSSRSSEMSPESSNFFSPRESTDF
    SSTSDAVASAGSGNRRDGPVPFAPVVSAGTSGRLLEADLDCDDGAGAV
    APGAGLIRSNCPASVAPQKHVVYKQARNCPAPASLKHODPDLDEEDDLVLTCE
    DDEDEMGSGSGVAGPEATLISIVDRTLLEPADKGEQVNFCESSNDGCPYEGCG
    AGAGLDGPGPTPSSYALTHPRPLPLDVPNGNQLVPPSSSQAPQGPENTAEHGA

```

```

gene
    <6121. .>14754
    /gene="tapasin"
    join(6121. .6206,6353. .6532,6731. .6991,11888. .12286,
    12486. .12827,13108. .13184,14618. .14754)
    /gene="tapasin"
    /product="tapasin"
    join(6170. .6206,6353. .6532,6731. .6991,11888. .12286,
    12486. .12827,13108. .13184,14618. .14629)
    /gene="tapasin"
    /note="intron-exon boundaries defined in relation to cDNA
    in GenBank Accession Numbers AF043943 and EST AA562898;
    EST W75528 contains an alternatively spliced form of this
    transcript"
    /codon_start=1
    /product="tapasin"
    /protein_id="AAC69893.1"
    /db_xref="GI:3811377"
    /translation="MKPLLLVAVAGLATVSVVSAGPEALIEGVFEDAGGGLSKK
    PATLLLRHGRPGPRPDLDKLYFKVDDPAGMLAARRRYPAGASAPHCMSRPIF
    PASAKWARSLSPEQNCPRALDMLVSVSTFSLSPROPELREPVITMATV
    VLVTLTNNPARVOLGKDAVLDLRPAVPSALESPSLDAGPPGLEMRORHGKCH
    LLIAATPGLAGMRPAQERKATAFAMDDDEPGWCGTGTMLPAVKRSGSCGLYATV
    HUPYIQGVSTLELYHKAPRVSLLTPAPVYMAAPGAPPELLCLASHFPBAGCYLWKE
    LRGGCGSSRVEGKVTWLSTRHSDGVSQGHQLDPPYAKOHGVHYVGRVYHSSL
    PASGSAADVTELVAAAWTIEVSEKATASLTIPRNSKRSQ"
    complement(8176. .8383)
    /rpt_family="B3"
    8545. .8670
    /rpt_family="B3"
    complement(9036. .9243)
    /rpt_family="MMB2"
    complement(9853. .9997)
    /rpt_family="MMB1"
    complement(10003. .10198)
    /rpt_family="MMB2"
    10003. .10014
    /gene="tapasin"
    /note="number of Ts unclear"
    10369. .10460
    /rpt_family="(GA)n"
    10454. .10474
    /gene="tapasin"
    /note="low quality data"
    complement(10836. .10943)
    /rpt_family="PBLD7"
    10973. .11020
    /rpt_family="MUSIDS"
    11215. .11328
    /rpt_family="MMB1F"
    11329. .11427
    /rpt_family="(GA)n"
    complement(11428. .11459)
    /rpt_family="(CA)n"
    11467. .11637
    /rpt_family="B3"
    complement(13691. .13830)
    /rpt_family="MMB1"
    complement(13859. .13993)
    /rpt_family="MWSQR2"
    complement(14790. .14932)
    /rpt_family="MMB1F"
    complement(14947. .15087)
    /rpt_family="MMB1"
    complement(15009. .15110)
    /rpt_family="MTE2"
    15133. .15243
    /rpt_family="MMB1"
    complement(15387. .15490)
    /rpt_family="UR1"

```

```

gene
16444..24873
/ gene="RLF"
/ note="Described as a Ras-associating protein in Genbank
Accession Number U54639"
mRNA
Join(16444..16522,16723..16919,18204..18287,18518..18696,
18799..18849,18932..19229,19487..19738,19829..19932,
20019..20103,20190..20258,20375..20449,21453..21485,
21612..21733,21838..21933,22023..22134,22218..22511,
23341..23455,23543..23754)
/ gene="RLF"
/ product="RalGDS-like factor"
CDS
Join(16764..16919,18204..18287,18518..18696,18799..18849,
18932..19229,19487..19738,19829..19932,20019..20103,
20190..20258,20375..20449,21453..21485,21612..21733,
21838..21933,22023..22134,22218..22511,23341..23455,
23543..23754)
Query Match 2.08: Score 183.6; DB 10; Length 273800;
Best Local Similarly 62.5%; Pred. No. 7,5e-42;
Matches 339; Conservative 0; Mismatches 194; Indels 9; Gaps 3;

QY 1853 GCGAGGAGAGAGCTGACGCTTCAGAGAGCAGCTTGTGCTTCAGAGAGCCTAGATT 1912
Db 107701 GCGCTGGAAGATGCGCCAGCGCTTAAGACACTGCGCTTGTGCGAGAGSA--TCCGTT 107644

QY 1913 CAGTTCCAGACTCATATGATGCTGCTCAGAGCCTCTGTAATCCAGTTCCAGAGGCTTC 1972
Db 107643 TGATTCCTAGATCCACAGAGTACTTCCAAACCATCTGTAACTCAGTGCAGGCGCTTG 107584

QY 1973 CACACCTCTTCTGCGCTCCAGAGCAGCAGCATATGATACAGACATATCATGCGAGCC 2032
Db 107583 GATGCCCTTTTCTGACCTCTACACTCTCCAGGAACAGCTGGTACACAGCTGTCAGGC 107524

QY 2033 AAAACACCCTATACACATATAATTAAGAACTTAAGAAAGTGCATGTGTTGTTAAAC 2092
Db 107523 AAAACACCAACACACAGGAATTAAAA---TAAGCTAGTTGAGTGGGTGATTAAAGGT 107467

QY 2093 ATTGTGCTTACATGCTGATTTGAAGACATGTACACGACACACACTGAGAGGATCTGG 2152
Db 107466 GAGGCGAGCGCGTCTCTGCAAGACTAAGTAGTAAGAACCTGCTTTAAAAACACCAAG 107407

QY 2153 GCGTGGAGAGATGGCTCAGGGTTAAGAGCACTAGCTCTTCCGAAGAGAGTCCCTGA 2212
Db 107406 GGTGGAGAGATGGCTCAGGGTTAAGAGCACTAGCTCTTCCAG---AGGCTCTTGA 107351

QY 2213 GTTCAATTCCTAGACACACATGCTGGCTTCACACACATTAATGATGTGACACCT 2272
Db 107350 GTTCAATTCCTAGACACACATGCTGGCTTCACACACATCTGTAGTGGATCTGATGCCCT 107291

QY 2273 CTTCGTGTCATCTGAAGACAGCTGACAGCTACAGTGTACTTATGATATTAATAATA 2332
Db 107290 TTCTGTGTCATCTGAAGACAGCTGACAGTGTACTTATGATATTAATAATA 107231

QY 2333 AATCTTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATGACAGT 2392
Db 107230 AAAAAAGATAAACAAACAAACCCCCCAAAACACAAACAAACACACTTTTAAATTGG 107171

QY 2393 GA 2394
Db 107170 GA 107169

```

```

RESULT 14
AC130278/c 182695 bp DNA linear HTG 09-AUG-2002
LOCUS Mus musculus clone RP24-393116, WORKING DRAFT SEQUENCE, 36
DEFINITION Unordered pieces.
AC130278
VERSION AC130278.1 GI:22165183
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 182695)
TITLE Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Mus musculus, clone RP24-393116
AUTHORS Unpublished
2 (bases 1 to 182695)
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Chepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karalas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schnpack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testfay,S., Theodorou,J., Tophan,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zalnoun,J.,
Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L25607
Center clone name: 393_1_16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168477 bases at least Q40
Consensus quality: 173980 bases at least Q30
Insert size: 178000; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-contigs
Quality coverage: 5.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 7168: contig of 7168 bp in length
* 7169 7268: gap of 100 bp
* 7269 7773: contig of 505 bp in length
* 7774 7873: gap of 100 bp
* 7874 8549: contig of 676 bp in length
* 8550 8649: gap of 100 bp
* 8650 9264: contig of 615 bp in length
* 9265 9364: gap of 100 bp
* 9365 10058: contig of 694 bp in length
* 10059 10158: gap of 100 bp
* 10159 10902: contig of 744 bp in length
* 10903 11002: gap of 100 bp
* 11003 11256: contig of 254 bp in length
* 11257 11356: gap of 100 bp
* 11357 12233: contig of 877 bp in length
* 12234 12333: gap of 100 bp

```



```
* 12334 13682: contig of 1349 bp in length
* 13683 13782: gap of 100 bp
* 13783 14874: contig of 1092 bp in length
* 14875 14974: gap of 100 bp
* 14975 16211: contig of 1237 bp in length
* 16212 16311: gap of 100 bp
* 16312 17319: contig of 1008 bp in length
* 17320 17419: gap of 100 bp
* 17420 19489: contig of 2070 bp in length
* 19490 19589: gap of 100 bp
* 19590 21253: contig of 1664 bp in length
* 21254 21353: gap of 100 bp
* 21354 22801: contig of 1448 bp in length
* 22802 22901: gap of 100 bp
* 22902 24454: contig of 1553 bp in length
* 24455 24534: gap of 100 bp
* 24535 25711: contig of 1157 bp in length
* 25712 25811: gap of 100 bp
* 25812 28015: contig of 2204 bp in length
* 28016 28115: gap of 100 bp
* 28116 29791: contig of 1676 bp in length
* 29792 29891: gap of 100 bp
* 29892 32114: contig of 2223 bp in length
* 32115 32214: gap of 100 bp
* 32215 32974: contig of 760 bp in length
* 32975 33074: gap of 100 bp
* 33075 33793: contig of 719 bp in length
* 33794 33893: gap of 100 bp
* 33894 37124: contig of 331 bp in length
* 37125 37224: gap of 100 bp
* 37225 40484: contig of 3260 bp in length
* 40485 40584: gap of 100 bp
* 40585 42891: contig of 2407 bp in length
* 42892 43091: gap of 100 bp
* 43092 48293: contig of 5202 bp in length
* 48294 48393: gap of 100 bp
* 48394 53122: contig of 4729 bp in length
* 53123 53222: gap of 100 bp
* 53223 75713: contig of 22493 bp in length
* 75716 75815: gap of 100 bp
* 75816 83666: contig of 7851 bp in length
* 83667 83766: gap of 100 bp
* 83767 92545: contig of 8779 bp in length
* 92546 92645: gap of 100 bp
* 92646 103351: contig of 10706 bp in length
* 103352 103451: gap of 100 bp
* 103452 118198: contig of 14747 bp in length
* 118199 118298: gap of 100 bp
* 118299 134605: contig of 16307 bp in length
* 134606 134705: gap of 100 bp
* 134706 151709: contig of 17004 bp in length
* 151710 151809: gap of 100 bp
* 151810 175524: contig of 23715 bp in length
* 175525 175624: gap of 100 bp
* 175625 182695: contig of 7071 bp in length.
```

## FEATURES

## Source

```
1. 182695
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-393i16"
/clone.lib="RPCI-24 Male Mouse BAC"
misc_feature
1. 7168
/note="assembly_fragment"
/clone.lib="RP24-393i16"
vector_side:left
misc_feature
7269.. 7773
/note="assembly_fragment"
7874.. 8549
/note="assembly_fragment"
misc_feature
8650.. 9264
/note="assembly_fragment"
misc_feature
9365.. 10058
/note="assembly_fragment"
```

```
misc_feature 10159..10902
/note="assembly_fragment"
misc_feature 11003..11256
/note="assembly_fragment"
misc_feature 11357..12233
/note="assembly_fragment"
misc_feature 12334..13682
/note="assembly_fragment"
misc_feature 13783..14874
/note="assembly_fragment"
misc_feature 14975..16211
/note="assembly_fragment"
misc_feature 16312..17319
/note="assembly_fragment"
misc_feature 17420..19489
/note="assembly_fragment"
misc_feature 19590..21253
/note="assembly_fragment"
misc_feature 21354..22801
/note="assembly_fragment"
misc_feature 22902..24454
/note="assembly_fragment"
misc_feature 24535..25711
/note="assembly_fragment"
misc_feature 25812..28015
/note="assembly_fragment"
misc_feature 28116..29791
/note="assembly_fragment"
misc_feature 29892..32114
/note="assembly_fragment"
misc_feature 32215..32974
/note="assembly_fragment"
misc_feature 33075..33793
/note="assembly_fragment"
misc_feature 33894..37124
/note="assembly_fragment"
misc_feature 37225..40484
/note="assembly_fragment"
misc_feature 40585..42991
/note="assembly_fragment"
misc_feature 43092..48293
/note="assembly_fragment"
misc_feature 48394..53122
/note="assembly_fragment"
```

## Query Match

Best Local Similarity 65.2%; Score 181.8; DB 2; Length 182695;

Matches 359; Conservative 0; Mismatches 137; Indels 55; Gaps 4;

```
OY 1797 ACATATAGTTACATGATGATTAATTAATTAATTTCTTTTAAAGGATATATGTTGGGA 1856
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171436 ACAGGTACTTACATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 171377
OY 1857 GGGAGAGATGGGTGCGCTTCCAGGACCTGCTGCTGCGAGGACCTAGATTCACT 1916
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171376 TAGAGAGATGAGTATAGTGTGTCAGAACACTGCTCTCTTCCAGGACCTTGTTCAAT 171317
OY 1917 TCCAGAGACTATATGTTGCTGTCACAGGCATCTGTAATTCAGTTCCAGAGGTTCCACA 1976
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171316 TCCAGACACATACATGTTGTCACAGCTGCTGAGTCTGATGTTCCAGGTTCCACACA 171257
OY 1977 CCTCTTTCGCTCCACAGCACCACAT----ACATAGTACACAGACATACATCAGGC 2032
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171256 GTCTTCTTGCGCTGCTGTTGATCCAGGTATGACATGATATACATAAACATGTCAGGT 171197
OY 2033 AAAACACCCATACACATTAATTAATTAAGAACTTAAGAGTGTCATGTTGTTAAAC 2092
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171196 AAAACACTCATCATATATAGATTAATAGATATGTTTAAAGG----- 171154
OY 2093 ATTGTCTTACACATGCTGATTGAAGACATGTACACACACACTGAAGAGGATCTGG 2152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171153 -----AAAGGAGAGAGAAAGAGAGACAGAAAAAGAAACAAAG 171116
```

Qy 2153 GCGTGAAGATGGCTCAGCGGTTAAGAGCAGTCGACTCTTCCGAGCAGAGGTCCTGA 2212  
||||| ||||||| ||| ||||||| ||||| |||||||  
Db 171115 GCGTGAAGATGGCTCAGCGGTTAAGAGCAGTCGACTCTTCC-----AAAGGTCCTGA 171060  
Qy 2213 GTTCAATCTAGCAGCACCATGCTGCTCACAACCATCCATATGAGATCTGCACCCCT 2272  
||||| ||||||| ||| ||||||| ||||||| ||||||| ||||||| |||  
Db 171059 GTTCAATCTAGCAGCACCATGCTGCTCACAACCATCCATATGAGATCTGCACCCCT 171000  
Qy 2273 CTCTCGTCGATCGAGCAGCTGACGAGCTACAGTCTAGATATGATATCTATAATA 2332  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 170999 CTCTCGTCGATCGAGCAGCTGACGAGCTACAGTCTAGATATGATATATAATA 170948  
Qy 2333 AATCTTTTTT 2343  
||||| |||||||  
Db 170947 AATCTTTTTT 170937

RESULT 15  
AC025964 218502 bp DNA linear HTG 09-MAY-2001  
LOCUS Mus musculus chromosome 11 clone RP23-136C7 map 11, WORKING DRAFT  
DEFINITION  
SEQUENCE 10 unordered pieces.  
AC025964  
AC025964.3 GI:13259480  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
VERSION  
KEYWORDS Mus musculus.  
SOURCE  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS 1 (bases 1 to 218502)  
Bliren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Boguslavsky,L., Boucknight,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choedel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczký,J.,  
Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Piere,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Tigililo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
DIRECT SUBMISSION

TITLE  
JOURNAL  
COMMENT  
Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 9, 2001 this sequence version replaced g1:880918.  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5729  
Center clone name: 136C-7  
----- Summary Statistics  
Sequencing vector: M13; M77815; 38% of reads  
Sequencing vector: Plasmid; n/a; 62% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 215619 bases at least Q40  
Consensus quality: 216900 bases at least Q30  
Consensus quality: 217250 bases at least Q20  
Insert size: 211000; agarose-fp  
Insert size: 217602; sum-of-ctnigs  
Quality coverage: 8.2 in Q20 bases; agarose-fp  
Quality coverage: 7.9 in Q20 ba.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 ctnigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the ctnigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 20205: contig of 20205 bp in length  
\* 20206 20305: gap of 100 bp  
\* 20306 25641: contig of 5336 bp in length  
\* 25642 25741: gap of 100 bp  
\* 25742 35304: contig of 9563 bp in length  
\* 35305 35404: gap of 100 bp  
\* 35405 45241: contig of 9837 bp in length  
\* 45242 45341: gap of 100 bp  
\* 45342 62833: contig of 17492 bp in length  
\* 62834 62933: gap of 100 bp  
\* 62934 110480: contig of 47547 bp in length  
\* 110481 110580: gap of 100 bp  
\* 110581 137309: contig of 26729 bp in length  
\* 137310 137409: gap of 100 bp  
\* 137410 166669: contig of 29260 bp in length  
\* 166670 166769: gap of 100 bp  
\* 166770 201541: contig of 34772 bp in length  
\* 201542 201641: gap of 100 bp  
\* 201642 218502: contig of 16861 bp in length.  
Location/Qualifiers  
1. 218502  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/map="11"  
/clone="RP23-136C7"  
/clone\_lib="RPC1-23 Female Mouse BAC"  
1. 20205  
/note="assembly-fragment  
clone\_end:SP6  
vector\_side:left"  
20306. 25641  
/note="assembly-fragment"  
25742. 35304  
/note="assembly-fragment"  
35405. 45241  
/note="assembly-fragment"  
45342. 62833  
/note="assembly-fragment"  
62934. 110480  
/note="assembly-fragment"  
110581. 137309  
/note="assembly-fragment"  
137410. 166669  
/note="assembly-fragment"  
166770. 201541  
/note="assembly-fragment"  
201642. 218502  
/note="assembly-fragment  
clone\_end:T7  
vector\_side:right"

BASE COUNT 57976 a 50523 c 50316 g 58775 t 912 others  
ORIGIN  
Query Match 2.0%; Score 179.2; DB 2; length 218502;  
Best Local Similarity 61.4%; Pred. No. 1.4e-40;



**THIS PAGE BLANK (USPTO)**



PN WO200034492-A1.  
XX 15-JUN-2000.  
XX  
PE 09-DEC-1998; 98WO-EP08009.  
XX  
PR 09-DEC-1998; 98WO-EP08009.  
XX  
PA (CNRS ) CENT NAT RECH SCI.  
PI (CURRI-) INST CURIE.  
XX  
PI Pinto D, Robline S, Jaisser F, Louvard D;  
DR WPI: 2000-423433/36.  
XX  
PT Novel nucleotide sequence derived from mouse villin gene for targeted  
PT expression of transgenes in immature and differentiated epithelial  
PT cells of intestine or urogenital tracts .  
XX  
PS Claim 3; Fig 6; 54pp; English.  
XX  
CC The present sequence comprises the murine villin gene regulatory  
CC region, the first Intron and part of the first two exons. It has been  
CC shown that this region directs the expression of the villin gene in the  
CC intestine and uro-genital tracts, and thus could be used in a fusion  
CC gene to direct expression of exogenous genes in these areas. This could  
CC be used, for example, to create a mouse model for colorectal cancer.  
XX  
SQ Sequence 8995 BP; 2275 A; 2105 C; 2258 G; 2357 T; 0 other:

Query Match 100.0%; Score 8995; DB 21; Length 8995;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTGCTGCACCAAGACACTGTGTCCTCCAGCACTGGGAGTGAGGAGAGAGGTCA 60  
DB 1 GATCTGCTGCACCAAGACACTGTGTCCTCCAGCACTGGGAGTGAGGAGAGGTCA 60  
QY 61 GAACCTTAAGGTCATCCCTGGTACATAGCAAGGTTCCAGCAGCTTCAGCTACATGAA 120  
DB 61 GAACCTTAAGGTCATCCCTGGTACATAGCAAGGTTCCAGCAGCTTCAGCTACATGAA 120  
QY 121 CCTTTGTTTGTGTTGTTGTTTAAAGCATTAAATATATACATAAGAGGTGG 180  
DB 121 CCTTTGTTTGTGTTGTTGTTTAAAGCATTAAATATATACATAAGAGGTGG 180  
QY 181 CAGTCGTGGCAGACACTTAATTCAGATATTCAGAGGCGAAGGAGGAGATCTCTGT 240  
DB 181 CAGTCGTGGCAGACACTTAATTCAGATATTCAGAGGCGAAGGAGGAGATCTCTGT 240  
QY 241 GAGTTCGAAGTCAGCTAGTCTGCAAAAGCTATGCCAGATGGGAGGCTACACAGAGA 300  
DB 241 GAGTTCGAAGTCAGCTAGTCTGCAAAAGCTATGCCAGATGGGAGGCTACACAGAGA 300  
QY 301 AACCTTCTCATTAACCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 360  
DB 301 AACCTTCTCATTAACCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 360  
QY 361 GTCCATTTCAGATGAGACATCTTAAGATGATTTCTTGACCCAGGTAACTATGTCA 420  
DB 361 GTCCATTTCAGATGAGACATCTTAAGATGATTTCTTGACCCAGGTAACTATGTCA 420  
QY 421 TGGGGAAGGAGGAGGAGCTGCTAGATTAAAGTGCTGAGGCGATGGCTATTCTCAA 480  
DB 421 TGGGGAAGGAGGAGGAGCTGCTAGATTAAAGTGCTGAGGCGATGGCTATTCTCAA 480  
QY 481 TTTGATTCCATATGAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGACTCTGGACT 540  
DB 481 TTTGATTCCATATGAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGACTCTGGACT 540  
QY 541 GAAGACGTGACGGCTTATTAACACGTGACACTTATTAACACTTATTAACACTGACACAG 600  
DB 541 GAAGACGTGACGGCTTATTAACACGTGACACTTATTAACACTTATTAACACTGACACAG 600

QY 601 CGTTTCAGGTTTGAAGATCACTTTCAAACCAAGAGAGAGAGAGTGTGCTGCTCAGC 660  
DB 601 CGTTTCAGGTTTGAAGATCACTTTCAAACCAAGAGAGAGAGTGTGCTGCTCAGC 660  
QY 661 GTAGCGACACACTGCTGCGAGAGAGTGTATTTTATGTAAGTACCTTACATATCTTT 720  
DB 661 GTAGCGACACACTGCTGCGAGAGAGTGTATTTTATGTAAGTACCTTACATATCTTT 720  
QY 721 GCACCTATCATACACAGTGTCAATGTGTACTCCCTATGTCACAGATGGCTGTAC 780  
DB 721 GCACCTATCATACACAGTGTCAATGTGTACTCCCTATGTCACAGATGGCTGTAC 780  
QY 781 CTGCTTTCTGCTTCCCATCTGCTGACATTTTCTCAACCAAGAAATTTAGAAATGTGGT 840  
DB 781 CTGCTTTCTGCTTCCCATCTGCTGACATTTTCTCAACCAAGAAATTTAGAAATGTGGT 840  
QY 841 ATTTATTTGTGTGCTGAGACACCATTCACAGGCTTTTACATTTTCAGGCATGTGTAC 900  
DB 841 ATTTATTTGTGTGCTGAGACACCATTCACAGGCTTTTACATTTTCAGGCATGTGTAC 900  
QY 901 TAACTGGGCTACTTCTCCAAACGGTTTGAACCATTTGTTTATATTTACTTATTTGTGT 960  
DB 901 TAACTGGGCTACTTCTCCAAACGGTTTGAACCATTTGTTTATATTTACTTATTTGTGT 960  
QY 961 GCATGAGGTAGGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
DB 961 GCATGAGGTAGGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
QY 1021 ATCATTCACAGATCCCAAGCAAGTGAAGTCCAGGCTGTGAATTTATGTGGACTG 1080  
DB 1021 ATCATTCACAGATCCCAAGCAAGTGAAGTCCAGGCTGTGAATTTATGTGGACTG 1080  
QY 1081 GGAGCCAAAGGCTGGTCTCTGCAAGAGGAGGCAAGTGGCTTACCATGGGAGCACCTGT 1140  
DB 1081 GGAGCCAAAGGCTGGTCTCTGCAAGAGGAGGCAAGTGGCTTACCATGGGAGCACCTGT 1140  
QY 1141 CTAGGCTTAAGTAAATCTTAAATATATATATATATATATATATATATATATATATAT 1200  
DB 1141 CTAGGCTTAAGTAAATCTTAAATATATATATATATATATATATATATATATATATAT 1200  
QY 1201 ACAAGCCCTTAATCCAGCACTTGAGAGGCTGAGTGAAGATTTATACACAGGCCAG 1260  
DB 1201 ACAAGCCCTTAATCCAGCACTTGAGAGGCTGAGTGAAGATTTATACACAGGCCAG 1260  
QY 1261 CTGGGGTGACAGCTGGGCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1320  
DB 1261 CTGGGGTGACAGCTGGGCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1320  
QY 1321 CTTGCGTGTATGTCCTGCAAGGGTGTCAAGATCCCTTGAAGCTGAGTTAAAGACAGTTG 1380  
DB 1321 CTTGCGTGTATGTCCTGCAAGGGTGTCAAGATCCCTTGAAGCTGAGTTAAAGACAGTTG 1380  
QY 1381 TGATCACGCTCCGTTACAGATCTGGAATTAACCCAGGTGCTCCCTAGAGAGCAGCC 1440  
DB 1381 TGATCACGCTCCGTTACAGATCTGGAATTAACCCAGGTGCTCCCTAGAGAGCAGCC 1440  
QY 1441 AGTGTCTTAACTTCTGAGCACCCCTCCAAACCTGCTTTTAAAGACTTTAACTCTTTG 1500  
DB 1441 AGTGTCTTAACTTCTGAGCACCCCTCCAAACCTGCTTTTAAAGACTTTAACTCTTTG 1500  
QY 1501 TGTAAATGTGGAACTGATGATTTGCACTTACCAAGTGTGCTGCGCTGAGCATCA 1560  
DB 1501 TGTAAATGTGGAACTGATGATTTGCACTTACCAAGTGTGCTGCGCTGAGCATCA 1560  
QY 1561 CTGAGCCCGTACCCACAGACTAGTGAATCACTTTAAGGGCAAACTTTAAACAATGACA 1620  
DB 1561 CTGAGCCCGTACCCACAGACTAGTGAATCACTTTAAGGGCAAACTTTAAACAATGACA 1620  
QY 1621 ATAGTTGATAGATTTGATATAGTCTGAGCTATTTGTTAGCGTACCTTTGCTGCTC 1680  
DB 1621 ATAGTTGATAGATTTGATATAGTCTGAGCTATTTGTTAGCGTACCTTTGCTGCTC 1680

Qy	1681	TTTACATGTGCTGTGGAGAGATGTGAAAAATGAAAGACTTGAGTCTTGCTCTGGAACCCACA	1740
Dp	1681	TTTACATGTGCTGTGGAGAGATGTGAAAAATGAAAGACTTGAGTCTTGCTCTGGAACCCACA	1740
Qy	1741	GAGCAGACGGCGAAGCCCACTCTGAAAGTGTGTTCTCTGAGCTTCACATACAACTTCACAT	1800
Dp	1741	GAGCAGACGGCGAAGCCCACTCTGAAAGTGTGTTCTCTGAGCTTCACATACAACTTCACAT	1800
Qy	1801	AATAGTTACATGTATTAATTAATTAATTAATTCCTTTTAAAAAGCTATATGTTTGGAGGGA	1860
Dp	1801	AATAGTTACATGTATTAATTAATTAATTAATTCCTTTTAAAAAGCTATATGTTTGGAGGGA	1860
Qy	1861	GAGATGGCTCAGCTTCAGAGACACTTGTCTCTTGACAGAGACCTTGATTCAGTTCC	1920
Dp	1861	GAGATGGCTCAGCTTCAGAGACACTTGTCTCTTGACAGAGACCTTGATTCAGTTCC	1920
Qy	1921	AGGACTATATGGTGGCTGCACACCCATCTGTAATCCATCTCCAGAGGTTCCACACCT	1980
Dp	1921	AGGACTATATGGTGGCTGCACACCCATCTGTAATCCATCTCCAGAGGTTCCACACCT	1980
Qy	1981	CTTCTGGCCCTCCACAGGCACACACATACATAGTACACAGACATACATGCAAGCCAAACACC	2040
Dp	1981	CTTCTGGCCCTCCACAGGCACACACATACATAGTACACAGACATACATGCAAGCCAAACACC	2040
Qy	2041	CATACACACATTAATTAATTAAGBAACCTTAAAGGTGCATGTGTTGTAACATTTGCT	2100
Dp	2041	CATACACACATTAATTAATTAAGBAACCTTAAAGGTGCATGTGTTGTAACATTTGCT	2100
Qy	2101	TACACATGTGATTTGAAGACATGTACACAGCACACACGTAAGAGGATCTGGGGCTGGAG	2160
Dp	2101	TACACATGTGATTTGAAGACATGTACACAGCACACACGTAAGAGGATCTGGGGCTGGAG	2160
Qy	2161	AGATGGCTCAGCGGTTAAAGAGCACTACTACTGCTCTTCCGAGGAAAGTCTGATGTTCAAT	2220
Dp	2161	AGATGGCTCAGCGGTTAAAGAGCACTACTACTGCTCTTCCGAGGAAAGTCTGATGTTCAAT	2220
Qy	2221	CTTAGCACACACATGGTGGCTCACAACCATCCATTAATGAGATCTGACACCTCTTGCT	2280
Dp	2221	CTTAGCACACACATGGTGGCTCACAACCATCCATTAATGAGATCTGACACCTCTTGCT	2280
Qy	2281	GCACTCAGACAGCTGCACAGGTACAGTGTACAGTGTACATATCTAATTAATTAATCTTTT	2340
Dp	2281	GCACTCAGACAGCTGCACAGGTACAGTGTACAGTGTACATATCTAATTAATTAATCTTTT	2340
Qy	2341	TTTTAAAAAATGAAGAGGATCTGAGACACACTCAAAAGAGATTATGACAGCTGACTCAG	2400
Dp	2341	TTTTAAAAAATGAAGAGGATCTGAGACACACTCAAAAGAGATTATGACAGCTGACTCAG	2400
Qy	2401	GGTGAATTATCTATCTCGAGCTTTTCTTCCCTTGCGCTTGCAACTGGTGGACAGCG	2460
Dp	2401	GGTGAATTATCTATCTCGAGCTTTTCTTCCCTTGCGCTTGCAACTGGTGGACAGCG	2460
Qy	2461	CCCCCTTTTCATTCACAAGAACGGGTCTCATATTATTTCGAACCAAAACAGCACTGCGAGT	2520
Dp	2461	CCCCCTTTTCATTCACAAGAACGGGTCTCATATTATTTCGAACCAAAACAGCACTGCGAGT	2520
Qy	2521	ATGTTTACTGTGCTGTGCTATATGAGCAAGGGCAAGGGCGGGGGGCGCACACACACACAC	2580
Dp	2521	ATGTTTACTGTGCTGTGCTATATGAGCAAGGGCAAGGGCGGGGGGCGCACACACACACAC	2580
Qy	2581	ACACACACACACACACACACACACACATTCATGAGTCTCCAGAGCTTGTGGGAAGTCA	2640
Dp	2581	ACACACACACACACACACACACACACATTCATGAGTCTCCAGAGCTTGTGGGAAGTCA	2640
Qy	2641	AGGAAGAGCTGCGCTCAAAACACATCTTATCTTTCCCTCTAAAGGAGACACAGATTC	2700
Dp	2641	AGGAAGAGCTGCGCTCAAAACACATCTTATCTTTCCCTCTAAAGGAGACACAGATTC	2700
Qy	2701	AAGGTGCGACAAAGTCTACAGGGGGGAGAGGCAAGGAGGGGGGAACAGGCCATGGTTCC	2760
Dp	2701	AAGGTGCGACAAAGTCTACAGGGGGGAGAGGCAAGGAGGGGGGAACAGGCCATGGTTCC	2760
Qy	2761	AGAGACCTACAGCAGAGGGGACGAAGGACAGATCCCAAGTCTCAGGGCAGGAGAGGTGAGG	2820

Db	2761	AGAGACCTACAGCGAGAGGGGACGAAAGGACGAGTCCCGAGGGGACGGGAGGGTGGAGG	2820
QY	2821	CCCTGTTCCGAGGAGAAGGACGAGCGGACGAAACAGGGTTCAAGGACACAGGTTTATGGCA	2880
Db	2821	CCCTGTTGTTCCGAGGAGAAGGACGAGCGGACGAAACAGGGTTCAAGGACACAGGTTTATGGCA	2880
QY	2881	GCTCATAAAAGTGGAGGTCGTGGCTCAGTCAGAAAGGAGGAAGGGAAGAGGCCCTGTGT	2940
Db	2881	GCTCATAAAAGTGGAGGTCGTGGCTCAGTCAGAAAGGAGGAAGGGAAGAGGCCCTGTGT	2940
QY	2941	GCCCACTGAGCGAGGGTCATGCTGAGTAGGAGAGATCTCAGGGGTGCCAGAGAGCCCCAC	3000
Db	2941	GCCCACTGAGCGAGGGTCATGCTGAGTAGGAGAGATCTCAGGGGTGCCAGAGAGCCCCAC	3000
QY	3001	CTGTCGTGCCCAAGGGAAGCCCAAGTGTGAACTCGTGGCTTGAGTTCAGCTA	3060
Db	3001	CTGTCGTGCCCAAGGGAAGCCCAAGTGTGAACTCGTGGCTTGAGTTCAGCTA	3060
QY	3061	CAAGACCCCAAGAGTCCCTACTCATCCCATTCACAGTGGCCCCCTCGCCCCGACACCCCA	3120
Db	3061	CAAGACCCCAAGAGTCCCTACTCATCCCATTCACAGTGGCCCCCTCGCCCCGACACCCCA	3120
QY	3121	CCCCGACTCCCGTGGCCACTTTCCTAGGGCTGGAGGGTGGCAGGCCCTGCTGGGGGTTGC	3180
Db	3121	CCCCGACTCCCGTGGCCACTTTCCTAGGGCTGGAGGGTGGCAGGCCCTGCTGGGGGTTGC	3180
QY	3181	CTACCTCGAGGTAGAGCCCAAGGTCCTAGGCGGAAGTGCACCCCACTCCCTGGAAGCGCAGA	3240
Db	3181	CTACCTCGAGGTAGAGCCCAAGGTCCTAGGCGGAAGTGCACCCCACTCCCTGAAGCTGCAGA	3240
QY	3241	GCCAAAGGCGGGGACACAGGACAGCTCAGGCTGTCAAGGCTGTGCTGGGCTCTAGAGTTCC	3300
Db	3241	GCCAAAGGCGGGGACACAGGACAGCTCAGGCTGTGCAAGGCTGTGCTGGGCTCTAGAGTTCC	3300
QY	3301	AGGAGACTGGGACCTTACTTCCCAACCCCACTCATCTCTCTGGGGCCCTATCTTCC	3360
Db	3301	AGGAGACTGGGACCTTACTTCCCAACCCCACTCATCTCTCTGGGGCCCTATCTTCC	3360
QY	3361	CTTATTATGTCGAAGCAAGTTCCTGGGGGGGGGGGGTGGTGTGAGACAAAGTGCTTGC	3420
Db	3361	CTTATTATGTCGAAGCAAGTTCCTGGGGGGGGGGGGTGGTGTGAGACAAAGTGCTTGC	3420
QY	3421	GTCCTCTGCACGACGCTTGGCACACACTCTCTAAGATCTCCAGAGTGTGGCTGCCTTTC	3480
Db	3421	GTCCTCTGCACGACGCTTGGCACACACTCTCTAAGATCTCCAGAGTGTGGCTGCCTTTC	3480
QY	3481	CAGACAGGTAAAGGCAATTTGGGTGGGACACATGTGTACACAGGTGTGGGAGGACAG	3540
Db	3481	CAGACAGGTAAAGGCAATTTGGGTGGGACACATGTGTACACAGGTGTGGGAGGACAG	3540
QY	3541	GSTCCTTGTCTCTCTCTGGAGGCTGTGGCTTCTGTAGACACTTGTAAAGTTTGGGGG	3600
Db	3541	GSTCCTTGTCTCTCTCTGGAGGCTGTGGCTTCTGTAGACACTTGTAAAGTTTGGGGG	3600
QY	3601	TGAGGTAAAGGTCCTGAACCTCTGAAGAAGCAAAACCCAGCAGGCTCTTGGGCT	3660
Db	3601	TGAGGTAAAGGTCCTGAACCTCTGAAGAAGCAAAACCCAGCAGGCTCTTGGGCT	3660
QY	3661	TCAATGAAGGAAGTTCACAGACCCCTTTCGTGTAGTACCTTGCTTGCATCTGTGAG	3720
Db	3661	TCAATGAAGGAAGTTCACAGACCCCTTTCGTGTAGTACCTTGCTTGCATCTGTGAG	3720
QY	3721	ATTCCCTGAGGACCAAGTGGCTCTCTGGGACTCAGATTTTCAAAATTAATCAGGACGT	3780
Db	3721	ATTCCCTGAGGACCAAGTGGCTCTCTGGGACTCAGATTTTCAAAATTAATCAGGACGT	3780
QY	3781	CCTGAGACTTGGACTCCGCTCTGTATTATCTCTCTGCTGGCTGCATTTCTGTGT	3840
Db	3781	CCTGAGACTTGGACTCCGCTCTGTATTATCTCTCTGCTGGCTGCATTTCTGTGT	3840
QY	3841	TCATGCTTACACATCTGAATGGTTTCTTGTGTGCACATTTCCCTGACACTCTGGGA	3900

```
Db 3841 TCATGTCCTTACACATCTGAATGTTCTTTGTGTCACCATTCCTCCGTACACTCTGGGA 3900
Qy 3901 GGTGCTATCCTTGGACATGATCTGGGATGTAAAGTCGACGACAGAGAGAGGGG 3960
Db 3901 GGTGCTATCCTTGGACATGATCTGGGATGTAAAGTCGACGACAGAGAGAGGGG 3960
Qy 3961 AGATGAGAGAGCTGTGCTTAGGCCCTATTAGGCTGGACATCACCCCTTCTAGAAAT 4020
Db 3961 AGATGAGAGAGCTGTGCTTAGGCCCTATTAGGCTGGACATCACCCCTTCTAGAAAT 4020
Qy 4021 GGGCCCTCCATTTTGGGTTACCATGATCTATTTATATCAGAGTGGGAGTAAAGCCA 4080
Db 4021 GGGCCCTCCATTTTGGGTTACCATGATCTATTTATATCAGAGTGGGAGTAAAGCCA 4080
Qy 4081 AACCTGCCAGAAATTGGGACTCAGTCAGACCAAGGTTATCTGCTCAGAAATCCCTGT 4140
Db 4081 AACCTGCCAGAAATTGGGACTCAGTCAGACCAAGGTTATCTGCTCAGAAATCCCTGT 4140
Qy 4141 TCACCTGAGGTTGGGAAATCTGCTTGGGGCTTCCAGGTCCTGGTTAGCAGAGGGT 4200
Db 4141 TCACCTGAGGTTGGGAAATCTGCTTGGGGCTTCCAGGTCCTGGTTAGCAGAGGGT 4200
Qy 4201 ATCCTTTGTATAGGCGATGACCTAGTCTATGCTGTACTATCTCTGCTCAGTTAAAG 4260
Db 4201 ATCCTTTGTATAGGCGATGACCTAGTCTATGCTGTACTATCTCTGCTCAGTTAAAG 4260
Qy 4261 CTGGAACCTAAACCCAGCGAGCGCCAGGATCTCTACAGTTGTACCCCAAGAACACA 4320
Db 4261 CTGGAACCTAAACCCAGCGAGCGCCAGGATCTCTACAGTTGTACCCCAAGAACACA 4320
Qy 4321 AGACAGTAGATATGCAAGATAGGTAGCTGGGAGAAAGAACTTAAACCCCCCAAG 4380
Db 4321 AGACAGTAGATATGCAAGATAGGTAGCTGGGAGAAAGAACTTAAACCCCCCAAG 4380
Qy 4381 GCCACAGGTTCCGTCCTCCAGTTCAACAATGCCAGTAGAGTCTAGCTATGGGCTG 4440
Db 4381 GCCACAGGTTCCGTCCTCCAGTTCAACAATGCCAGTAGAGTCTAGCTATGGGCTG 4440
Qy 4441 TGAGTTGGTACTACAGCATGAGTGTATGTCATGCTGAGTGTATATATCTGAGCAC 4500
Db 4441 TGAGTTGGTACTACAGCATGAGTGTATGTCATGCTGAGTGTATATATCTGAGCAC 4500
Qy 4501 TTGGGAGGCTGAAGCAGAGAGATTTCTATATGTTTGAAGCCCTGAGCTATAGACGA 4560
Db 4501 TTGGGAGGCTGAAGCAGAGAGATTTCTATATGTTTGAAGCCCTGAGCTATAGACGA 4560
Qy 4561 GACTTTGCTTTAAAGAAAAATGAAGCCACAGTGTGGACACAGCCCTTAAATCCCA 4620
Db 4561 GACTTTGCTTTAAAGAAAAATGAAGCCACAGTGTGGACACAGCCCTTAAATCCCA 4620
Qy 4621 GCACCTGGAGGAGAGAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGTGTATAGAGT 4680
Db 4621 GCACCTGGAGGAGAGAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGTGTATAGAGT 4680
Qy 4681 GAGTTCAGAGCAGCCAGGCTACACAGAGAAACCTGTTTGAAGAACAGAAAAAACA 4740
Db 4681 GAGTTCAGAGCAGCCAGGCTACACAGAGAAACCTGTTTGAAGAACAGAAAAAACA 4740
Qy 4741 AACAAAAAACAACAAACAAACCCAAACCCAAACCTCATCTCATCTCTC 4800
Db 4741 AACAAAAAACAACAAACAAACCCAAACCCAAACCTCATCTCATCTCTC 4800
Qy 4801 TAGGCTGTGCTGTAGTGTAGAGTTTGGGACTTACAGTTATATATATATAGGCC 4860
Db 4801 TAGGCTGTGCTGTAGTGTAGAGTTTGGGACTTACAGTTATATATATATAGGCC 4860
Qy 4861 TTTTATACACTGTGTAGAGAGAGAAAGTTTACAGTCTGGACACAGTGGACCTGAGA 4920
Db 4861 TTTTATACACTGTGTAGAGAGAGAAAGTTTACAGTCTGGACACAGTGGACCTGAGA 4920
Qy 4921 AAGTACTCTTGGCAGCCCAAAATTTCTGGAGAGGCTTCTTGAGAGAAAGTGTCCGAT 4980
Db 4921 AAGTACTCTTGGCAGCCCAAAATTTCTGGAGAGGCTTCTTGAGAGAAAGTGTCCGAT 4980
Qy 4981 CAGACTACTGTTCTAGAAAGCAGAGAGGTTGGAAGAAATGTTGTGACAGACAGTT 5040
Db 4981 CAGACTACTGTTCTAGAAAGCAGAGAGGTTGGAAGAAATGTTGTGACAGACAGTT 5040
Qy 5041 GGAACAGAAAGCAGAGAGGGGGAGGATCCAAAGATTCTGACATGTAGCTATTTGGT 5100
Db 5041 GGAACAGAAAGCAGAGAGGGGGAGGATCCAAAGATTCTGACATGTAGCTATTTGGT 5100
Qy 5101 TCTCTGGGTGACAGTGTCTCCCGAGGATAGGGCTTAGAAAGGGGACCAAGGCTAGACC 5160
Db 5101 TCTCTGGGTGACAGTGTCTCCCGAGGATAGGGCTTAGAAAGGGGACCAAGGCTAGACC 5160
Qy 5161 AATGAGTTCAAGTTGAGGAGACATCAGCCAGGCTCTTCTGCGCAAGCTAAAGAAAG 5220
Db 5161 AATGAGTTCAAGTTGAGGAGACATCAGCCAGGCTCTTCTGCGCAAGCTAAAGAAAG 5220
Qy 5221 AGAGCCCTTAACCTCCCTGAAGTTTAAAGGAGACAGAGAGCTGAGAGATCTTCTA 5280
Db 5221 AGAGCCCTTAACCTCCCTGAAGTTTAAAGGAGACAGAGAGCTGAGAGATCTTCTA 5280
Qy 5281 GGGTGAAGGAGAGGTATCTGCTGTGACAAATAGGCTAGAGCAGAGAGTGGACAG 5340
Db 5281 GGGTGAAGGAGAGGTATCTGCTGTGACAAATAGGCTAGAGCAGAGAGTGGACAG 5340
Qy 5341 TTACCCCTCAGAACAGACCATCCCTTGGCTTAAGAGAGGCTGGCCCTTCTGTT 5400
Db 5341 TTACCCCTCAGAACAGACCATCCCTTGGCTTAAGAGAGGCTGGCCCTTCTGTT 5400
Qy 5401 AAGATCTTACTTTTCTTACAGAGAGACAGACAGCTTTGTCCCTCCTGTTGCTAA 5460
Db 5401 AAGATCTTACTTTTCTTACAGAGAGACAGACAGCTTTGTCCCTCCTGTTGCTAA 5460
Qy 5461 TTAACACCCCTGTGTAACTATGATTATTTACTGTAGTTGTGCTCAGACAGTCCA 5520
Db 5461 TTAACACCCCTGTGTAACTATGATTATTTACTGTAGTTGTGCTCAGACAGTCCA 5520
Qy 5521 TCTGTAGACCTCTGCTCTTAACATCACAAGGATGGCCACATTTCTCACCCAGAAAG 5580
Db 5521 TCTGTAGACCTCTGCTCTTAACATCACAAGGATGGCCACATTTCTCACCCAGAAAG 5580
Qy 5581 TGCAGAGAGAGCTTGAAGAAAGGTTAAACAGTAAACAAAGTGCCAGATTAACAAAA 5640
Db 5581 TGCAGAGAGAGCTTGAAGAAAGGTTAAACAGTAAACAAAGTGCCAGATTAACAAAA 5640
Qy 5641 ACTACTATCCTTTGTACCAAAATGTTTGTGTAACCAAGAGGGGTGTGAGGTAT 5700
Db 5641 ACTACTATCCTTTGTACCAAAATGTTTGTGTAACCAAGAGGGGTGTGAGGTAT 5700
Qy 5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
Db 5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
Qy 5761 CTTGGGGACTTTTCACTGCTAAAGAAATATCTGATTTGGCGCCCAAGAGGGGTAT 5820
Db 5761 CTTGGGGACTTTTCACTGCTAAAGAAATATCTGATTTGGCGCCCAAGAGGGGTAT 5820
Qy 5821 TGGGAGAGTCAAGGCTTTCGAAACACAGTAAGCTCCCAAGTGAATTGTGGCTGAA 5880
Db 5821 TGGGAGAGTCAAGGCTTTCGAAACACAGTAAGCTCCCAAGTGAATTGTGGCTGAA 5880
Qy 5881 TCACCAAGGGGAGGCTGTATCAGAGTGAACAATCAAGATTAAGCACCTGTGGG 5940
Db 5881 TCACCAAGGGGAGGCTGTATCAGAGTGAACAATCAAGATTAAGCACCTGTGGG 5940
Qy 5941 GCTCAGAGAGGAGTTTACAGAGGTAAAGGCCAAGCCATTTATTTATCAAGACATGAC 6000
Db 5941 GCTCAGAGAGGAGTTTACAGAGGTAAAGGCCAAGCCATTTATTTATCAAGACATGAC 6000
Qy 6001 TCAAAATCAAGTCAAGAGAGATTAAGTGAAGATAGGGCTGTCAAGTGTGGGACACC 6060
Db 6001 TCAAAATCAAGTCAAGAGAGATTAAGTGAAGATAGGGCTGTCAAGTGTGGGACACC 6060
```



Qy	6061	TGACCTTGGACACTTATTAGTCACCTAGGCCCAAGAGCAGTCACAGAGGGTGACTGGGTCTTA	6120
Dp	6061	TGACCTTGGACACTTATTAGTCACCTAGGCCCAAGAGCAGTCACAGAGGGTGACTGGGTCTTA	6120
Qy	6121	CTCAGCTTGGAGAGCGGCACCTGGAGAAATGGGTGACCTCCATCTCATGATGAGAGGGCTGAG	6180
Dp	6121	CTCAGCTTGGAGAGCGGCACCTGGAGAAATGGGTGACCTCCATCTCATGATGAGAGGGCTGAG	6180
Qy	6181	CACCACCAGGTAGCAGTGTGTCCCTGTGTCTCATAGCCAGATTCCAGGCGAGTGTTCACAA	6240
Dp	6181	CACCACCAGGTAGCAGTGTGTCCCTGTGTCTCATAGCCAGATTCCAGGCGAGTGTTCACAA	6240
Qy	6241	GACTAAGGACTCATCTCTGCTGGAAACAAGATGCCAGCCCTAAGCCCCATTTTGGTCT	6300
Dp	6241	GACTAAGGACTCATCTCTGCTGGAAACAAGATGCCAGCCCTAAGCCCCATTTTGGTCT	6300
Qy	6301	AATTAATATCAAGAACCCCTGGGGATGTGAGGCTCTGAGCGACAGCAGACTTTTAAAAAGTC	6360
Dp	6301	AATTAATATCAAGAACCCCTGGGGATGTGAGGCTCTGAGCGACAGCAGACTTTTAAAAAGTC	6360
Qy	6361	CCAGAGTATTTCTGATCAGCAGCTGGAACAACACAGAGCTTCAGAGTTCAAACAGAAAGAGC	6420
Dp	6361	CCAGAGTATTTCTGATCAGCAGCTGGAACAACACAGCTTCAGAGTTCAAACAGAAAGAGC	6420
Qy	6421	AAAGCTAAGGAAGCTTGGGATGGGAGCCCTTCCTCCAGGCCAGTAGATGAGAGCTGGT	6480
Dp	6421	AAAGCTAAGGAAGCTTGGGATGGGAGCCCTTCCTCCAGGCCAGTAGATGAGAGCTGGT	6480
Qy	6481	AGCAGTGTGGGACAGCTTCTCTGCTGTGCATATAGCTATTCATCCACTCATCCATCAT	6540
Dp	6481	AGCAGTGTGGGACAGCTTCTCTGCTGTGCATATAGCTATTCATCCACTCATCCATCAT	6540
Qy	6541	ACACCCACCCCATCATTTTATGCAACCCATCTTCGATCCATCATATACAGCATACCCAC	6600
Dp	6541	ACACCCACCCCATCATTTTATGCAACCCATCTTCGATCCATCATATACAGCATACCCAC	6600
Qy	6601	CCACGCAATCAGCAACCTTCCTTTTCCTCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT	6660
Dp	6601	CCACGCAATCAGCAACCTTCCTTTTCCTCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT	6660
Qy	6661	CATTATATCCAAAGACAACCTGGTATGTGTAATAATGGGAAATTTAATTTTATGA	6720
Dp	6661	CATTATATCCAAAGACAACCTGGTATGTGTAATAATGGGAAATTTAATTTTATGA	6720
Qy	6721	AGCTCTTTATTTAGCACTGATTTGGCATGTATGAGGACAGGTATATACACAGCAGCAGCTG	6780
Dp	6721	AGCTCTTTATTTAGCACTGATTTGGCATGTATGAGGACAGGTATATACACAGCAGCAGCTG	6780
Qy	6781	TGGCAATCGGAGAAAGTTTGGGTGTTTCTTCCTCCACCGCTGTGGGTTCGGGA	6840
Dp	6781	TGGCAATCGGAGAAAGTTTGGGTGTTTCTTCCTCCACCGCTGTGGGTTCGGGA	6840
Qy	6841	TTGGAATCAAAATTTATGGGGCTGGTGGCAAGTCTTTTACACCCAGCCATTTTGTGTACA	6900
Dp	6841	TTGGAATCAAAATTTATGGGGCTGGTGGCAAGTCTTTTACACCCAGCCATTTTGTGTACA	6900
Qy	6901	CATCATTTATTTAGAAAGCATTTATGTATGTCAGAGCTGGGCTTAAGCTTGCTATGTGCG	6960
Dp	6901	CATCATTTATTTAGAAAGCATTTATGTATGTCAGAGCTGGGCTTAAGCTTGCTATGTGCG	6960
Qy	6961	CCACGATGACCTTTAAGCTCTGCTCTTCACAGCTTCACCCGAGTGTAGGTTTACAGCT	7020
Dp	6961	CCACGATGACCTTTAAGCTCTGCTCTTCACAGCTTCACCCGAGTGTAGGTTTACAGCT	7020
Qy	7021	GTTCAACTGGTGAATGCTTTAATCCACAGCTCTGTGTGGGGGGGGGGGGAGCGGATC	7080
Dp	7021	GTTCAACTGGTGAATGCTTTAATCCACAGCTCTGTGTGGGGGGGGGGGGAGCGGATC	7080
Qy	7081	CCTGAGTTGAGGCAAGTTTGTGTCACAGAGTTTCAGATACCTGGGCTATACAGGGA	7140
Dp	7081	CCTGAGTTGAGGCAAGTTTGTGTCACAGAGTTTCAGATACCTGGGCTATACAGGGA	7140
Qy	7141	ACCCTATCCCAACAAACAACAACAAATAATTTCTGTGCAATTAATACAGAGA	7200

Dd	7141	ACCCATATCCCAACAAACAACAAACAAACAAAATAATTCTGTGCATTAATACACAGAGA	7200
Oy	7201	TTAGAGATATTTAGTAGGAGTATAGGGCTGTCAGAGAGCATGCTTTCCTTGTATY	7260
Dd	7201	TTAAGCATTATAGTACGGAGTAGTACGGCTCGTACGGAAGTCATGCTTCTTTTGATY	7260
Oy	7261	ATAATAGTAAAGTACTCACAAAGATGCATATCTATCTATCTATCTATCTATCTATYC	7320
Dd	7261	ATAATAGTAAAGTACTCACAAAGATGCATATCTATCTATCTATCTATCTATCTATYC	7320
Oy	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	7380
Dd	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	7380
Oy	7381	CTGCTTTGACATCGAATAGCTCCATTTCTGGGGCAACTCTAACCCATAGTTGGGGTTY	7440
Dd	7381	CTGCTTTGACATCTGAATAGCTCTCTATTTCTGGGGCAACTCTAACCCATAGTTGGGGTTY	7440
Oy	7441	ACCAACACCAGACATTTATTTATTTTGTATTTATTTATTTATCTAGAGCTCAGSGTG	7500
Dd	7441	ACCAACACCAGACATTTATTTATTTTGTATTTATTTATTTATCTAGAGCTCAGSGTG	7500
Oy	7501	GGACTCAGGGCTCTTGTCATGCTAATGCAAGAGCTCTGCCACAGAGCTGAGCTCCAGTCC	7560
Dd	7501	GGACTCAGGGCTCTTGTCATGCTAATGCAAGAGCTCTGCCACAGAGCTGAGCTCCAGTCC	7560
Oy	7561	CCAATTTGTTTACAGSTGACCTGTGACAGTTGTCATTTTGGAGGCTATGTAGCTCTCTC	7620
Dd	7561	CCAATTTGTTTACAGSTGACCTGTGACAGTTGTCATTTTGGAGGCTATGTAGCTCTCTC	7620
Oy	7621	CACCTCCACATWCCAGACATTTCTGGTTCATCCAGTGGGGGGGCAACTCTGTGTCACCA	7680
Dd	7621	CACCTCCACATWCCAGACATTTCTGGTTCATCCAGTGGGGGGGCAACTCTGTGTCACCA	7680
Oy	7681	GTGCCCTGTTTCCCTGTCTTCAGACCTPACATATTTGCTGTGGAACAGTTCAATGTAATG	7740
Dd	7681	GTGCCCTGTTTCCCTGTCTTCAGACCTPACATATTTGCTGTGGAACAGTTCAATGTAATG	7740
Oy	7741	GGATGGCTTCCCTGTGATTCCTTTATGGGCGGCCCTTATCTTATGACACAGTTTGTGTG	7800
Dd	7741	GGATGGCTTCCCTGTGATTCCTTTATGGGCGGCCCTTATCTTATGACACAGTTTGTGTG	7800
Oy	7801	GGCATTGTGTCACGTGCTATCTATCTATCTATCATCATCTTATAGGCTTAATAGTGTCCYT	7860
Dd	7801	GGCATTGTGTCACGTGCTATCTATCTATCTATCATCATCTTATAGGCTTAATAGTGTCCYT	7860
Oy	7861	TGTGTGATTAACACACTTCTGTTTCATTTACATGATGAAATTTGTGCCCCACCCCAC	7920
Dd	7861	TGTGTGATTAACACACTTCTGTTTCATTTACATGATGAAATTTGTGCCCCACCCCAC	7920
Oy	7921	CCTTTTTTTTTTATTTAGACAAAGGCTTCTGTCTATCTTGAACATCTGGCTGTCT	7980
Dd	7921	CCTTTTTTTTTTATTTAGACAAAGGCTTCTGTCTATCTTGAACATCTGGCTGTCT	7980
Oy	7981	GAGCTCACTCTGTAGACACAGGCTGTAGGCTGCTCTCACTTTGACACTCTGTGAAC	8040
Dd	7981	GAGCTCACTCTGTAGACACAGGCTGTAGGCTGCTCTCACTTTGACACTCTGTGAAC	8040
Oy	8041	AGAGTACCCATGAACCTTCAAGACAAATTTTCTGTGTTTTGTTTTTATCATTTGTGTG	8100
Dd	8041	AGAGTACCCATGAACCTTCAAGACAAATTTTCTGTGTTTTGTTTTTATCATTTGTGTG	8100
Oy	8101	TGTATGSGTGTATATGTGCAATGTTTGTGCTTCAAGTGTCTCAACATGTGTAGCTGTGTG	8160
Dd	8101	TGTATGSGTGTATATGTGCAATGTTTGTGCTTCAAGTGTCTCAACATGTGTAGCTGTGTG	8160
Oy	8161	TGGGACAGAGAAACAAACCGATGTGCCATCTCTAGATFACATGCACTTGTATATATGTA	8220
Dd	8161	TGGGACAGAGAAACAAACCGATGTGCCATCTCTAGATFACATGCACTTGTATATATGTA	8220
Oy	8221	TGTATATATGATGTTTATTTATAGTGTGCCAAGTATGCAAGTATTTTGTGGAGTTTTCAC	8280
Dd	8221	TGTATATATGATGTTTATTTATAGTGTGCCAAGTATGCAAGTATTTTGTGGAGTTTTCAC	8280

```

Db 8221 TGTATTATGATGTTTATTTAGTGTGCCCAAGTATGACGATTTTCTGTGAGTTTTCAC 8280
Oy 8281 CTTCCCTTGTGGGCTCTCCGATTAACATCAGCTCTCGGGCTAGTAGCAATGCTTCA 8340
Db 8281 CTTCCCTTGTGGGCTCTCCGATTAACATCAGCTCTCGGGCTAGTAGCAATGCTTCA 8340
Oy 8341 CTGATGAGCCATCTGCTGCGCCCTGCTGCGACCTCTCTTATTTTCCAGATGGACTA 8400
Db 8341 CTGATGAGCCATCTGCTGCGCCCTGCTGCGACCTCTCTTATTTTCCAGATGGACTA 8400
Oy 8401 CGCAGCTGAGCTGAGCTTAAAGCTCAACCAAGTATCAGAGTGGCTAGGAGGAGACTCAG 8460
Db 8401 CGCAGCTGAGCTGAGCTTAAAGCTCAACCAAGTATCAGAGTGGCTAGGAGGAGACTCAG 8460
Oy 8461 GGATATGCTGGCCCTCTGCTCTCAGACAGTCTAGATTAACAGCATATCAGCTGTGGAAG 8520
Db 8461 GGATATGCTGGCCCTCTGCTCTCAGACAGTCTAGATTAACAGCATATCAGCTGTGGAAG 8520
Oy 8521 ATTTTAACTGAATCTCTGAGGATAGAGCAGGACCTCTACCAATGAGAGGTTCTTTTGT 8580
Db 8521 ATTTTAACTGAATCTCTGAGGATAGAGCAGGACCTCTACCAATGAGAGGTTCTTTTGT 8580
Oy 8581 GTTTGGTTGGTTCTCTCTGATATAGATCAAGGAGCTGAAATAGTGTAGCTGGCTAC 8640
Db 8581 GTTTGGTTGGTTCTCTCTGATATAGATCAAGGAGCTGAAATAGTGTAGCTGGCTAC 8640
Oy 8641 ATTAACATCTTGTCTCAAAAAAGCCTATAGAGTAGAGGAGGCTAAAGAAAGGCTT 8700
Db 8641 ATTAACATCTTGTCTCAAAAAAGCCTATAGAGTAGAGGAGGCTAAAGAAAGGCTT 8700
Oy 8701 AAGCGGCTGTGATAGACACAGAGATAGCCTGACTATATAGCAAGAAGCTTTGTTCAAA 8760
Db 8701 AAGCGGCTGTGATAGACACAGAGATAGCCTGACTATATAGCAAGAAGCTTTGTTCAAA 8760
Oy 8761 ACATGGAGGAGGAGGCTATGTTTAAAGTCTGGGCTGTCTAAACAGCACTAAAGGAGCCAA 8820
Db 8761 ACATGGAGGAGGAGGCTATGTTTAAAGTCTGGGCTGTCTAAACAGCACTAAAGGAGCCAA 8820
Oy 8821 TGTAGACATTTGATAGAAAGATCATCATCAAAAGCGGGTGGGAGGAGTAGAGTTGG 8880
Db 8821 TGTAGACATTTGATAGAAAGATCATCATCAAAAGCGGGTGGGAGGAGTAGAGTTGG 8880
Oy 8881 ACTACAGTGTGCAAGACCCCATAGGAAGCCAGTTTCCCTTCTCTGCGGCTCAAGC 8940
Db 8881 ACTACAGTGTGCAAGACCCCATAGGAAGCCAGTTTCCCTTCTCTGCGGCTCAAGC 8940
Oy 8941 CTGGCTGAGGCGCACAGCTCTCATGAGCTTCTCTGAGGCTGCTCAGCAGT 8995
Db 8941 CTGGCTGAGGCGCACAGCTCTCATGAGCTTCTCTGAGGCTGCTCAGCAGT 8995

```

RESULT 2  
AAAA9605 standard; DNA: 8993 BP.

```

ID AAAA9605 standard; DNA: 8993 BP.
XX
XX AAAA9605;
AC
XX 19-DEC-2000 (first entry)
DT
XX
XX Mouse villin gene regulatory elements and partial sequence.
DE
XX Mouse; villin; intestinal epithelial cell;
KM uro-genital tract epithelial cell; tumour; ds.
XX
OS Mus sp.
XX
XX key Location/Qualifiers
FH key 3442
FT misc-signal /tag= a
FT exon /note= "transcriptional start site"
FT 3442..3487
FT /tag= b
FT /number= 1

```

```

FT Intron 3488..8981
FT /tag= c
FT /number= 1
FT exon 8982..8995
FT /tag= d
FT /number= 2
FT misc-signal 8992
FT /tag= e
FT /note= "translational start site"
XX
XX W0200034493-A2.
XX
XX I5-JUN-2000.
XX
XX 09-DEC-1999; 99MO-EP09782.
XX
XX 09-DEC-1998; 98MO-EP08009.
XX
XX (CNRS ) CENT NAT RECH SCI.
XX (CURIE-) INST CURIE.
XX
XX Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J;
XX WPI: 2000-423434/36.
XX
XX Novel nucleotide sequence derived from mouse villin gene for targeted
XX expression of transgenes in immature and differentiated epithelial
XX cells of intestine or urogenital tracts -
XX
XX Claim 3: Fig 6; 52pp; English.
XX
XX The present sequence consists of the regulatory region and the first
XX exon and intron of the murine villin gene. This gene is expressed in the
XX epithelial cells of the intestine and uro-genital tracts. Its promoter
XX sequence can be used in the targeted expression of exogenous genes in
XX these places, which may, for example, be useful in the treatment of
XX tumours.
XX
XX Sequence 8993 BP; 2274 A; 2105 C; 2257 G; 2355 T; 2 other;
XX
XX Query Match 99.7%; Score 8969.6; DB 21; Length 8993;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 8991; Conservative 1; Mismatches 1; Indels 2; Gaps 2;

```

|||||  
Db 421 TGGGAAAGGGAGTGGAGCTGTCTAGATTAAAGTGTGAGGAGTCTATTCTCAA 480  
Oy 481 TTTGATTTCCATATGAAAAGCGTGAATAGGCCCAAGAGAAGTGGAACTGTGGACTGTGGACT 540  
Db 481 TTTGATTTCCATATGAAAAGCGTGAATAGGCCCAAGAGAAGTGGAACTGTGGACTGTGGACT 540  
Oy 541 GAAGACGTGACGGCGCTTTAAACACTGGACCTTAAACACTTAAACACTGGACAGG 600  
Db 541 GAAGACGTGACGGCGCTTTAAACACTGGACCTTAAACACTTAAACACTGGACAGG 600  
Oy 601 CGTTGAGGTTTGAAAGATCACTTTCAAAACACAGAAAGAAAGTGGTCCGCTCCGTCACG 660  
Db 601 CGTTGAGGTTTGAAAGATCACTTTCAAAACACAGAAAGAAAGTGGTCCGCTCCGTCACG 660  
Oy 661 GTAGCGAGCACTGGCTGCGAAGAGATGATATTAGTAAAGCTACCTTCACAAATCTTT 720  
Db 661 GTAGCGAGCACTGGCTGCGAAGAGATGATATTAGTAAAGCTACCTTCACAAATCTTT 720  
Oy 721 GCACATTACATACAGGTGCAAAAGTGTACCTCCAGTCCAGTCCAGTGGCTGTAC 780  
Db 721 GCACATTACATACAGGTGCAAAAGTGTACCTCCAGTCCAGTCCAGTGGCTGTAC 780  
Oy 781 CTCGTTTCGCTTTCCCATCTGTGGTGAACATTTGTGAGAACACAGAAATTTAGAATGTGGGT 840  
Db 781 CTCGTTTCGCTTTCCCATCTGTGGTGAACATTTGTGAGAACACAGAAATTTAGAATGTGGGT 840  
Oy 841 ATTTATTTGTGTGCTGAGAGACCAATCCAGGGCTTTTCACATTCAGGACATGCTTTAC 900  
Db 841 ATTTATTTGTGTGCTGAGAGACCAATCCAGGGCTTTTCACATTCAGGACATGCTTTAC 900  
Oy 901 TAACGGGGCTACTTCCAAAGGTTTGAACCATTTGTTTTATTTACTTATTTGTGT 960  
Db 901 TAACGGGGCTACTTCCAAAGGTTTGAACCATTTGTTTTATTTACTTATTTGTGT 960  
Oy 961 GCATGAGGTAGGACATGATATACGTATAGAGTCAATGATGCTGCTACCTCCAAA 1020  
Db 961 GCATGAGGTAGGACATGATATACGTATAGAGTCAATGATGCTGCTACCTCCAAA 1020  
Oy 1021 ATCATTTGCAAGATCCCAGACAGTGAAGTACCCGAGCGTTGTAGTTTGTATGTGGAGCTG 1080  
Db 1021 ATCATTTGCAAGATCCCAGACAGTGAAGTACCCGAGCGTTGTAGTTTGTATGTGGAGCTG 1080  
Oy 1081 GGAGCGAAGGCTGGGTTCTGTCAGAGAGAGCGAGTGGCTTAACCATGGGACAGCTCT 1140  
Db 1081 GGAGCGAAGGCTGGGTTCTGTCAGAGAGAGCGAGTGGCTTAACCATGGGACAGCTCT 1140  
Oy 1141 CTAGGCTTAAGGTAATCTTTAGTTTTTAAATAATATATTTCTAGCGCGGTGTGGTGC 1200  
Db 1141 CTAGGCTTAAGGTAATCTTTAGTTTTTAAATAATATATTTCTAGCGCGGTGTGGTGC 1200  
Oy 1201 ACAGCGCTTTAATCCCAGACACTTGAAGAGCTGAGAGTGTAGAAATTAACACAGGCCAG 1260  
Db 1201 ACAGCGCTTTAATCCCAGACACTTGAAGAGCTGAGAGTGTAGAAATTAACACAGGCCAG 1260  
Oy 1261 CTGGGGTGCAGAGCTTGGGCGCTTTTTTTTTTTTTTTCTTATATGAGCACTGGTGTCTTA 1320  
Db 1261 CTGGGGTGCAGAGCTTGGGCGCTTTTTTTTTTTTTTTCTTATATGAGCACTGGTGTCTTA 1320  
Oy 1321 CCTGGGTATATGTCGGTCAAGAGGTGTAGATCCCTGGAGCTGAGTTAAAGACAGTTG 1380  
Db 1321 CCTGGGTATATGTCGGTCAAGAGGTGTAGATCCCTGGAGCTGAGTTAAAGACAGTTG 1380  
Oy 1381 TGATACCGCTGCGCTTACAGATGCTGGAATTTGAACCCAGAGTGTCCCTAGAGACAGCG 1440  
Db 1381 TGATACCGCTGCGCTTACAGATGCTGGAATTTGAACCCAGAGTGTCCCTAGAGACAGCG 1440  
Oy 1441 AGTGTCTTAATCTTGAAGCCACCCCTCAACCTGCTTTTGAAGACCTTAAACCTTTTG 1500  
Db 1441 AGTGTCTTAATCTTGAAGCCACCCCTCAACCTGCTTTTGAAGACCTTAAACCTTTTG 1500  
Oy 1501 TGTAAATGTGGAACTGAGTGAATCTTGACATTAACAAAGTGTGCTGGCTGTAGCATCA 1560  
Db 1501 TGTAAATGTGGAACTGAGTGAATCTTGACATTAACAAAGTGTGCTGGCTGTAGCATCA 1560

Db 1501 TGTAAATGTGGAACTGAGTGAATCTTGACATTAACAAAGTGTGCTGGCTGTAGCATCA 1560  
Oy 1561 CTGAGCGCGCTACCCACAGCACTAGTGTGATACAGTTTAAAGGCAACACTTAACATGACA 1620  
Db 1561 CTGAGCGCGCTACCCACAGCACTAGTGTGATACAGTTTAAAGGCAACACTTAACATGACA 1620  
Oy 1621 ATAGTTGGATAGATTTTGAATTTAGCTTACCTGAGCTATTTGGTTTGGCGGACCTTTGCTGCC 1680  
Db 1621 ATAGTTGGATAGATTTTGAATTTAGCTTACCTGAGCTATTTGGTTTGGCGGACCTTTGCTGCC 1680  
Oy 1681 TTAGCATGTGCTGTGAGAGATAGAAAATGAAGACTTGTAGTCTAGCTGGAACCCACA 1740  
Db 1681 TTAGCATGTGCTGTGAGAGATAGAAAATGAAGACTTGTAGTCTAGCTGGAACCCACA 1740  
Oy 1741 GAGGCAAGCGAGAACCCACTCCTGAAAGTTGTTCTGTGAGCTTCACATTAACATTCACAT 1800  
Db 1741 GAGGCAAGCGAGAACCCACTCCTGAAAGTTGTTCTGTGAGCTTCACATTAACATTCACAT 1800  
Oy 1801 AATAGTTACAAATGATTAATTAATTAATTAATTTCTTTTAAAGGTAATGTTGGGAGGA 1860  
Db 1801 AATAGTTACAAATGATTAATTAATTAATTAATTTCTTTTAAAGGTAATGTTGGGAGGA 1860  
Oy 1861 GAGATGGCTCAAGCTTCCAGAGACACTTGGCTGCTTTGACAGAGACCTAGATTCACTTCCC 1920  
Db 1861 GAGATGGCTCAAGCTTCCAGAGACACTTGGCTGCTTTGACAGAGACCTAGATTCACTTCCC 1920  
Oy 1921 AGGACTCATATGTGTGGCTCACAGCCATCTGTAAATCCATTCAGAGGTTCCACACCT 1980  
Db 1921 AGGACTCATATGTGTGGCTCACAGCCATCTGTAAATCCATTCAGAGGTTCCACACCT 1980  
Oy 1981 CTTTGGGCTCCAGAGGACCCACATATAGTATAGACAGACATATATGCGAGCAAAACAC 2040  
Db 1981 CTTTGGGCTCCAGAGGACCCACATATAGTATAGACAGACATATATGCGAGCAAAACAC 2040  
Oy 2041 CATACACACATTAATTAATTAAGAAACTTAAAGGTGCATGTGTTGTAAACATTTGCT 2100  
Db 2041 CATACACACATTAATTAATTAAGAAACTTAAAGGTGCATGTGTTGTAAACATTTGCT 2100  
Oy 2101 TACACATGCTGATTTGAAGACATGTACAAAGCACACATGAAGAGGATGTGGGGCTGGAG 2160  
Db 2101 TACACATGCTGATTTGAAGACATGTACAAAGCACACATGAAGAGGATGTGGGGCTGGAG 2160  
Oy 2161 AGATGGCTCAGCGGTTTGAAGAGCACTGAGCTGCTTCCGAAAGAGGATCTGTGATTCAAAT 2220  
Db 2161 AGATGGCTCAGCGGTTTGAAGAGCACTGAGCTGCTTCCGAAAGAGGATCTGTGATTCAAAT 2220  
Oy 2221 CCTAGCAACACATGTTGGCTCACAAACATCCATATATGATCTGACACCTCTTCTGCT 2280  
Db 2221 CCTAGCAACACATGTTGGCTCACAAACATCCATATATGATCTGACACCTCTTCTGCT 2280  
Oy 2281 GCATCTGAAGACAGCTGCGAGAGCTACAGTGTACTTATGATATCTAATTAATTAATCTTTT 2340  
Db 2281 GCATCTGAAGACAGCTGCGAGAGCTACAGTGTACTTATGATATCTAATTAATTAATCTTTT 2340  
Oy 2341 TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATAGAGCACTGACTCAG 2400  
Db 2341 TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATAGAGCACTGACTCAG 2400  
Oy 2401 GGTGATTAATCTATCTGAGAGTTTCTCTTCCGCTTGGCTTGCACACTGGGTGAGACAG 2460  
Db 2401 GGTGATTAATCTATCTGAGAGTTTCTCTTCCGCTTGGCTTGCACACTGGGTGAGACAG 2460  
Oy 2461 CCCCCTTTTCAATTCACAAAGAGGAGGCTGATCAATTAATTTTGAACAAAGCACCTGCACT 2520  
Db 2461 CCCCCTTTTCAATTCACAAAGAGGAGGCTGATCAATTAATTTTGAACAAAGCACCTGCACT 2520  
Oy 2521 ATGTTTACGCTCTGCTGACTATAGACAGCGGACAGCGCGGCGGCACACACACACAC 2580  
Db 2521 ATGTTTACGCTCTGCTGACTATAGACAGCGGACAGCGCGGCGGCACACACACACAC 2580  
Oy 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGGTCA 2640  
Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGGTCA 2640

OY	264.1	AGAGAGGCTGCCCTCAAACAGCATCTTATCTTTCCCTCCTTAAAGGAGACCAGCATTC	2700
Db	264.1	AGAAAGAGGCTGCCCTCAAACAGCATCTTATCTTTCCCTCCTTAAAGGAGACCAGCATTC	2700
OY	270.1	AAGGCGCAGAAAGTCTACAGGGGGCAGAGGCCAGGGAGGGGGACACAGGCCATGGTTTCC	2760
Db	270.1	AAGTGGCAGAAAGTCTACAGGGGGCAGAGGCCAGGGAGGGGGAGCAGGCCATGGTTTCC	2760
OY	276.1	AGAGACCTTACAGCAGAGGGGACGACAGGACAGATCCCAAGTCCAGGGCAGGGAGGTGGAGG	2820
Db	276.1	AGAGACCTTACAGCAGAGGGGACGACAGGACAGATCCCAAGTCCAGGGCAGGGAGGTGGAGG	2820
OY	282.1	CCCTTTGTTCCGAGAGAAAGCAGCGGCGACAAAGGTTCAAAGCCACAGTTTATGGCA	2880
Db	282.1	CCCTTTGTTCCGAGAGAAAGCAGCGGCGACAAAGGTTCAAAGCCACAGTTTATGGCA	2880
OY	288.1	GCTCATTAAGAGTGGAGGTCTGGCTCAGTCAGAAAGAGGAAGAAAGGAAAGGCCCTTGT	2940
Db	288.1	GCTCATTAAGAGTGGAGGTCTGGCTCAGTCAGAAAGAGGAAGAAAGGAAAGGCCCTTGT	2940
OY	294.1	GCCCACTGAGCGAGGGTCACTGCTAGTAGAGAGATCTGCAAGGGGTGGCAGAGACCCAC	3000
Db	294.1	GCCCACTGAGCGAGGGTCACTGCTAGTAGAGAGATCTGCAAGGGGTGGCAGAGACCCAC	3000
OY	300.1	CTGTCTGTCCTCCCAAGGGAACCCCAAGTGAACCTTGCGCTTGAGTTCACACTA	3060
Db	300.1	CTGTCTGTCCTCCCAAGGGAACCCCAAGTGAACCTTGCGCTTGAGTTCACACTA	3060
OY	306.1	CAAAACCCCAAGAGTCCCTACTCCATCCCATCAGTCCGCCCCCTCCCGCCGACACCCCA	3120
Db	306.1	CAAAACCCCAAGAGTCCCTACTCCATCCCATCAGTCCGCCCCCTCCCGCCGACACCCCA	3120
OY	312.1	CCCCGAGTCCCGGCGCACTTCTCTAGAGGCTGGAGGGTGGCAGGCCCTGGTGGGGTTGC	3180
Db	312.1	CCCCGAGTCCCGGCGCACTTCTCTAGAGGCTGGAGGGTGGCAGGCCCTGGTGGGGTTGC	3180
OY	318.1	CTACCTGCAGGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCCTGGAAGCTGAGA	3240
Db	318.1	CTACCTGCAGGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCCTGGAAGCTGAGA	3240
OY	324.1	GCCAAAGGGCGGGGACACAGGCAAGCTAGAGGCTGTGCTGGGCTCTAGGTTCC	3300
Db	324.1	GCCAAAGGGCGGGGACACAGGCAAGCTAGAGGCTGTGCTGGGCTCTAGGTTCC	3300
OY	330.1	AGGAGCCGGGACACCTACTTCCCAACCCCCCATCATTTCTCTGAGGAGCCCTATCTTCC	3360
Db	330.1	AGGAGCCGGGACACCTACTTCCCAACCCCCCATCATTTCTCTGAGGAGCCCTATCTTCC	3360
OY	336.1	CTTATATGCTGAAGGAATTCTGAGGGGGGGGGGCTGTGTGAGAGCAAAAGTCTTTCG	3420
Db	336.1	CTTATATGCTGAAGGAATTCTGAGGGGGGGGGGCTGTGTGAGAGCAAAAGTCTTTCG	3420
OY	342.1	GTTCTCTGCAAGCCAGCTTGGCCACAATCTCTTAAGATCTTCCAGGTGGTGGCTCTTC	3480
Db	342.1	GTTCTCTGCAAGCCAGCTTGGCCACAATCTCTTAAGATCTTCCAGGTGGTGGCTCTTC	3480
OY	348.1	CAGACAGGTAAAGCAATTTGGGTGGGGAACATGTGTGACACAGGCTGTGTGAAGGGGACAG	3540
Db	348.1	CAGACAGGTAAAGCAATTTGGGTGGGGAACATGTGTGACACAGGCTGTGTGAAGGGGACAG	3540
OY	354.1	GGTCTTGTCTTCTCTGCGCAGCCTTGCTTCTGTAGCACCCTTGATTAAGTTTGGGG	3600
Db	354.1	GGTCTTGTCTTCTCTGCGCAGCCTTGCTTCTGTAGCACCCTTGATTAAGTTTGGGG	3600
OY	360.1	TGAGGTTAAGGTGCTCTGAACCTCTGAAGAAGCAAGAGCCAGCAGGCTGTCTTGGGCT	3660
Db	360.1	TGAGGTTAAGGTGCTCTGAACCTCTGAAGAAGCAAGAGCCAGCAGGCTGTCTTGGGCT	3660
OY	366.1	TCAAATGAAGGAAGTTACACAGACCCCTTTCCTGTAAAGTACACTTCGCTTAACTGTGTAG	3720
Db	366.1	TCAAATGAAGGAAGTTACACAGACCCCTTTCCTGTAAAGTACACTTCGCTTAACTGTGTAG	3720

QY	3721	ATTCCCTGGGACCAAGTGGCTCCTGGGACTAGATTTCTACAATTAAATACAGACAGT	3780
Db	3721	ATTCCCTGGGACCAAGTGGCTCCTGGGACTAGATTTCTACAATTAAATACAGACAGT	3780
QY	3781	CCTGAGACTTGGACCTCGCTCGTATTATACATCTCTCTGCGCTCATATTTCTGTGT	3840
Db	3781	CCTGAGACTTGGACCTCGCTCGTATTATACATCTCTCTGCGCTCATATTTCTGTGT	3840
QY	3841	TCATGCTCTACACATCTGAAATGTTTTTGTGTGTCACATTCCCTGACACTCTGTGGG	3900
Db	3841	TCATGCTCTACACATCTGAAATGTTTTTGTGTGTCACATTCCCTGACACTCTGTGGG	3900
QY	3901	GGTGCTATCTCTGGCAGCATATCTATCTCGGGATGTAACTCTACCCACAGAGAGAGGGG	3960
Db	3901	GGTGCTATCTCTGGCAGCATATCTATCTCGGGATGTAACTCTACCCACAGAGAGAGGGG	3960
QY	3961	AGAGTCAAGAGCTGTGTCTCTAGAGCCCTATTAGCCGTGAGCATCACCCCTTCTCTAGAA	4020
Db	3961	AGAGTCAAGAGCTGTGTCTCTAGAGCCCTATTAGCCGTGAGCATCACCCCTTCTCTAGAA	4020
QY	4021	GGCCCTCCATTTTTGGGTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGCA	4080
Db	4021	GGCCCTCCATTTTTGGGTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGCA	4080
QY	4081	AACCTGCCACAGAGTTTGGGACTCTACCTACAGACCAGAGTTATCTGCTCAAGAAATCCCCCTG	4140
Db	4081	AACCTGCCACAGAGTTTGGGACTCTACCTACAGACCAGAGTTATCTGCTCAAGAAATCCCCCTG	4140
QY	4141	TCACCTAGAGTGGGAGAAATCTGCTGGGGCTCCAGGCTTGTTAGAGAGAGGT	4200
Db	4141	TCACCTAGAGTGGGAGAAATCTGCTGGGGCTCCAGGCTTGTTAGAGAGAGGT	4200
QY	4201	ATTCCTTTGTATAGGCGATGACCTAGCTATATGAGTTACTACATTCCTCTCCAGTTAAAG	4260
Db	4201	ATTCCTTTGTATAGGCGATGACCTAGCTATATGAGTTACTACATTCCTCTCCAGTTAAAG	4260
QY	4261	CTGSACTAAAAACCCACGCGCAGCCAGAGTTCTCTACAGTTGTACCCCAAGAACACA	4320
Db	4261	CTGSACTAAAAACCCACGCGCAGCCAGAGTTCTCTACAGTTGTACCCCAAGAACACA	4320
QY	4321	AGACAGTAGATATGCAAGGATAGTAGTGCGGGAGAGAAAGCTTAAACCCCCCAAG	4380
Db	4321	AGACAGTAGATATGCAAGGATAGTAGTGCGGGAGAGAAAGCTTAAACCCCCCAAG	4380
QY	4381	GCCACAGATTCGTTCCGTCCCTAGTTTCAATATGCCAGTATAGGTCTACTATATGGGCTG	4440
Db	4381	GCCACAGATTCGTTCCGTCCCTAGTTTCAATATGCCAGTATAGGTCTACTATATGGGCTG	4440
QY	4441	TGAGTTGGTAGTACACAGCATGAGTAGTGTTCATGTGTAGTGTATTAATCTGACAC	4500
Db	4441	TGAGTTGGTAGTACACAGCATGAGTAGTGTTCATGTGTAGTGTATTAATCTGACAC	4500
QY	4501	TTGGGAGGCTGAAAGCAGAGAGATTGCTATATGTTTGAGGCCAGCTTGAGCTATAGAGGA	4560
Db	4501	TTGGGAGGCTGAAAGCAGAGAGATTGCTATATGTTTGAGGCCAGCTTGAGCTATAGAGGA	4560
QY	4561	GACTTTGCTTTTAAAGAAAAATGAAGGCCAGAGCTGTGGCAGACAGCCCTTTAATCCCA	4620
Db	4561	GACTTTGCTTTTAAAGAAAAATGAAGGCCAGAGCTGTGGCAGACAGCCCTTTAATCCCA	4620
QY	4621	GCACCTTGGGAGGAGAGCAGGCGATTTCTAGTTCAAGGCCAGCCGTGCTATAGAGT	4680
Db	4621	GCACCTTGGGAGGAGAGCAGGCGATTTCTAGTTCAAGGCCAGCCGTGCTATAGAGT	4680
QY	4681	GAGTTCCAGAGCAGCCAGGGCTTCAAGAGAGAAACCTGTTTTGAAAAACAGAAAAACAA	4740
Db	4681	GAGTTCCAGAGCAGCCAGGGCTTCAAGAGAGAAACCTGTTTTGAAAAACAGAAAAACAA	4740
QY	4741	AACAAAACAAAACAAAACAAAACCCCAAAACCCAAACCCAAACCTCTCATCTCTCATCTTC	4800
Db	4741	AACAAAACAAAACAAAACAAAACCCCAAAACCCAAACCCAAACCTCTCATCTCTCATCTTC	4800
QY	4801	TAGAGCTGTGCTGTAGTGTAGAGTTGGGAGCTTCAGACTATATATTAATAGGCC	4860

|||||  
Db 4801 TAGCTGTGTCTGTAGGTGTAGAGTTTGGGACTTCAGACTATATATATAATAGGCC 4860  
Oy 4861 TTTTATACATGTGTACAGACGAGAAAGTTTCACTCTGGGACACAGTGGACCTTGAGA 4920  
Db 4861 TTTTATACATGTGTACAGACGAGAAAGTTTCACTCTGGGACACAGTGGACCTTGAGA 4920  
Oy 4921 AAGTCTCTTCCAGCCCAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT 4980  
Db 4921 AAGTCTCTTCCAGCCCAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT 4980  
Oy 4981 CAGACTACTGTCTAGAAAGGCGAGAGAGAGGTTGGAAGAAATGTGTGACAGACAGT 5040  
Db 4981 CAGACTACTGTCTAGAAAGGCGAGAGAGGTTGGAAGAAATGTGTGACAGACAGT 5040  
Oy 5041 GGAACAGAGAGCAGAGAGGGGAGGCATCCAAATTTCTGAACATGTAGCTGACTTTGGT 5100  
Db 5041 GGAACAGAGAGCAGAGAGGGGAGGCATCCAAATTTCTGAACATGTAGCTGACTTTGGT 5100  
Oy 5101 TCTCTGGGTGACAGTGTCCCCAGAGGATAGGGCTGTAGAAAGGGGACCAGGGGTGAGCC 5160  
Db 5101 TCTCTGGGTGACAGTGTCCCCAGAGGATAGGGCTGTAGAAAGGGGACCAGGGGTGAGCC 5160  
Oy 5161 AATGAGTTCAAGTTAGGGGACACATCCAGCCAGGGGTCTTGCTGGCAAGCTTAAAGATG 5220  
Db 5161 AATGAGTTCAAGTTAGGGGACACATCCAGCCAGGGGTCTTGCTGGCAAGCTTAAAGATG 5220  
Oy 5221 AGAGCCCTCTAACCTCCCTGGAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTTCTA 5280  
Db 5221 AGAGCCCTCTAACCTCCCTGGAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTTCTA 5280  
Oy 5281 GGGTGAAGAGAGATATCTGCTCTGACCAACATGGCTAGGACAGAGAGATTTGGACAG 5340  
Db 5281 GGGTGAAGAGAGATATCTGCTCTGACCAACATGGCTAGGACAGAGAGATTTGGACAG 5340  
Oy 5341 TTACCCCTCAGAACCCAGCCATCCCTCTTGGCTTAAGAGAGCTGGGCCCTTCTGT 5400  
Db 5341 TTACCCCTCAGAACCCAGCCATCCCTCTTGGCTTAAGAGAGCTGGGCCCTTCTGT 5400  
Oy 5401 AAGATCTTACTTTCTTCTCAGAGAGAGCAGAACCTTTTCCCTCTCTTGTGTCAA 5460  
Db 5401 AAGATCTTACTTTCTTCTCAGAGAGAGCAGAACCTTTTCCCTCTCTTGTGTCAA 5460  
Oy 5461 TAAACACCCCTGTGTAAACATTATTATTACTGTGTCAGTTTCTCCAGAGACGTCA 5520  
Db 5461 TAAACACCCCTGTGTAAACATTATTATTACTGTGTCAGTTTCTCCAGAGACGTCA 5520  
Oy 5521 TCTGTAGACCTCTGCTCTACTACCAAGGATATGGCCACATTCCTCACCCAGAGAG 5580  
Db 5521 TCTGTAGACCTCTGCTCTACTACTACCAAGGATATGGCCACATTCCTCACCCAGAGAG 5580  
Oy 5581 TGCACAAAGAGCCTTAGAGAAAGGTTAAACAGTAACAAAGATGGCCAGATTAACAA 5640  
Db 5581 TGCACAAAGAGCCTTAGAGAAAGGTTAAACAGTAACAAAGATGGCCAGATTAACAA 5640  
Oy 5641 ACTACTATCTTGTACCAAAATTTGTTTGTGAACAGAGAGGGGTGTGAGTGTAT 5700  
Db 5641 ACTACTATCTTGTACCAAAATTTGTTTGTGAACAGAGAGGGGTGTGAGTGTAT 5700  
Oy 5701 GT 5760  
Db 5701 GT 5760  
Oy 5761 CTGGGGGACTTTTATGCTAAAGAAATATCTATATTTGGCCGCCATGCCAACAGGGAT 5820  
Db 5761 CTGGGGGACTTTTATGCTAAAGAAATATCTATATTTGGCCGCCATGCCAACAGGGAT 5820  
Oy 5821 TGGGAGAGTCAAGCTTGTGCAACACAGTAGTCCCAAGATGGATTGGGGCTGAA 5880  
Db 5821 TGGGAGAGTCAAGCTTGTGCAACACAGTAGTCCCAAGATGGATTGGGGCTGAA 5880  
Oy 5881 TCACCAAGGGGAGGCTATCAGATGAGACAGAACATCAAGATTAAGCACCCTGTGGG 5940  
|||||

Db 5881 TCACCAAGGGGAGGCTATCAGATGAGACAGAACATCAAGATTAAGCACCCTGTGGG 5940  
Oy 5941 GCTCAGAAAGAGGAGTTTACAAAGAGGTAAAGCCAAAGCATTATTATTCAAAGACATGAC 6000  
Db 5941 GCTCAGAAAGAGGAGTTTACAAAGAGGTAAAGCCAAAGCATTATTATTCAAAGACATGAC 6000  
Oy 6001 TCAAAATCAAAAGTGCAGAGAGATAGCTGAGAGATGGGGCTGTCAAGTGTGGGACAC 6060  
Db 6001 TCAAAATCAAAAGTGCAGAGAGATAGCTGAGAGATGGGGCTGTCAAGTGTGGGACAC 6060  
Oy 6061 TGACCTTCACCTTATTATGACATAGGCCAAGAGGACATCCAGAGAGGTACTGGTCTTA 6120  
Db 6061 TGACCTTCACCTTATTATGACATAGGCCAAGAGGACATCCAGAGAGGTACTGGTCTTA 6120  
Oy 6121 CTCACCTTGGAGCAGGACAGCTGTGCAATGTGGTACCTCATCTGTATGAGAGGGCTGAG 6180  
Db 6121 CTCACCTTGGAGCAGGACAGCTGTGCAATGTGGTACCTCATCTGTATGAGAGGGCTGAG 6180  
Oy 6181 CACACACAGGTACAGTGTCCCTGTGTCTCATGACCAGGATTCCTGGCCAGTTTCAAG 6240  
Db 6181 CACACACAGGTACAGTGTCCCTGTGTCTCATGACCAGGATTCCTGGCCAGTTTCAAG 6240  
Oy 6241 GACTAAGGACTATCTGTGTGGAAACAAAGTATCCAAAGCCCTAAGCCCATTTTGTCT 6300  
Db 6241 GACTAAGGACTATCTGTGTGGAAACAAAGTATCCAAAGCCCTAAGCCCATTTTGTCT 6300  
Oy 6301 AATTAAATCAGAACCCCTGGGGATGAGGCTGTGAGACAGACAGACTTTTAAAAAGCTC 6359  
Db 6301 AATTAAATCAGAACCCCTGGGGATGAGGCTGTGAGACAGACAGACTTTTAAAAAGCTC 6359  
Oy 6361 CCAGGTATTCGATCAGCAGCTGGAACAAACACAGCTACAGTTTCAAAACAGAAAGAGC 6420  
Db 6361 CCAGGTATTCGATCAGCAGCTGGAACAAACACAGCTACAGTTTCAAAACAGAAAGAGC 6420  
Oy 6421 AAAGCTAGGGAAGCTTGGGATGGGAGGCTTCTTCAGAGCCAGTAGATGAGGCTGTT 6480  
Db 6421 AAAGCTAGGGAAGCTTGGGATGGGAGGCTTCTTCAGAGCCAGTAGATGAGGCTGTT 6480  
Oy 6481 AGCAGTGTGGAGGCTTCTCTGCTGTGATATAGCTATTCATCCAGTACATCCAT 6540  
Db 6481 AGCAGTGTGGAGGCTTCTCTGCTGTGATATAGCTATTCATCCAGTACATCCAT 6540  
Oy 6541 ACACCCACCCATCCATTAATGACACCAATCTTCATATCCATATCCAGCTACCCAC 6600  
Db 6541 ACACCCACCCATCCATTAATGACACCAATCTTCATATCCATATCCAGCTACCCAC 6600  
Oy 6601 CCAGCATTCATCAAAACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Db 6601 CCAGCATTCATCAAAACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Oy 6661 CATTTATCCACAGAGAACTGTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6720  
Db 6661 CATTTATCCACAGAGAACTGTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6720  
Oy 6721 AGCTGTGTGATGTGACTGATTTGTGCATGTATGAGACAGGTACATACACACACAGCTG 6780  
Db 6721 AGCTGTGTGATGTGACTGATTTGTGCATGTATGAGACAGGTACATACACACACAGCTG 6780  
Oy 6781 TGGCAATGGAGAAAGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6840  
Db 6781 TGGCAATGGAGAAAGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6840  
Oy 6841 TTGAACCTCAAAATTAATCGGGCTGGGCAAGTCTTTTACACACGAGCCATTTTGTCTGACA 6900  
Db 6841 TTGAACCTCAAAATTAATCGGGCTGGGCAAGTCTTTTACACACGAGCCATTTTGTCTGACA 6900  
Oy 6901 CATCATTTATTTAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGAT 6960  
Db 6901 CATCATTTATTTAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGAT 6960  
Oy 6961 CCACGGATGACCTTTTAACTCTGCTCTTCCAGCCTTCACCCGAGGTAGTTTACAGGT 7020  
Db 6961 CCACGGATGACCTTTTAACTCTGCTCTTCCAGCCTTCACCCGAGGTAGTTTACAGGT 7020

OY	7021	GTTTCAACTGCTGAATGCCCTTTAAATCCACAGACTCTGTGGGGGGGGGGAGCGGATTC	7080
Dd	7020	GTTTCAACTGCTGAATGCCCTTTAAATCCACAGACTCTGTGGGGGGGGGGAGCGGATTC	7079
OY	7081	CCTGAGTGCGAGCCGATTGGCTCTACAGATTTCAGATTAACCTGGGGGCATATCAGGGAA	7140
Dd	7080	CCTGAGTGCGAGCCGATTGGCTCTACAGATTTCAGATTAACCTGGGGGCATATCAGGGAA	7139
OY	7141	ACCCTATCCCAAACAACAACAACAACAAAAAATAATTTCTGTGCATTAATCACAGAGA	7200
Dd	7140	ACCCTATCCCAAACAACAACAACAACAAAAAATAATTTCTGTGCATTAATCACAGAGA	7199
OY	7201	TTAGAGGATATTAGTAGGCTAGTAGGGCTGTGAGGGAGATCATCTTCTTTGTATT	7260
Dd	7200	TTAGAGGATATTAGTAGGCTAGTAGGGCTGTGAGGGAGATCATCTTCTTTGTATT	7259
OY	7261	ATAATAGTAATACACACAAGATGATTAATCTATCTATCTATCTATCTATCTATC	7320
Dd	7260	ATAATAGTAATAGTACACACAAGATGATTAATCTATCTATCTATCTATCTATCTATC	7319
OY	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC	7380
Dd	7320	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC	7379
OY	7381	CTGCTTTGACTCTGAATGCTCTATTCTTGGGCTCAACTCTTCACCCCAGTGTGGGTTT	7440
Dd	7380	CTGCTTTGACTCTGAATGCTCTATTCTTGGGCTCAACTCTTCACCCCAGTGTGGGTTT	7439
OY	7441	ACCAACACCCACACANTTTTTATTTTTATTTTTATTTATATATCAGAGGACGAGGTTG	7500
Dd	7440	ACCAACACCCACACANTTTTTATTTTTATTTTTATTTATTTATTTATTAATCAGAGGACGAGGTTG	7499
OY	7501	GGACTCAGGGCTTGTGATGCTTAAGCAAGCTCTCTGCCACAGAGCTGGAGCTCCAGTCC	7560
Dd	7500	GGACTCAGGGCTTGTGATGCTTAAGCAAGCTCTCTGCCACAGAGCTGGAGCTCCAGTCC	7559
OY	7561	CCATTTTGTTCAGAGTGACCTGTGACAGTTGTTCATATTGGCAGCGCTATATGTAAGCTCTTC	7620
Dd	7560	CCATTTTGTTCAGAGTGACCTGTGACAGTTGTTCATATTGGCAGCGCTATATGTAAGCTCTTC	7619
OY	7621	CACCTCCCAAGTTCACAGCACTTTCTGGTCATCCAGGGGGGCACTCTGTCCTACCA	7680
Dd	7620	CACCTCCCAAGTTCACAGCACTTTCTGGTCATCCAGGGGGGCACTCTGTCCTACCA	7679
OY	7681	GTCGCCGTGTTCCCTGTCTTTCAGACCTACATATTTGCCGTGTGAAAGTTTCATGTAAATG	7740
Dd	7680	GTCGCCGTGTTCCCTGTCTTTCAGACCTACATATTTGCCGTGTGAAAGTTTCATGTAAATG	7739
OY	7741	GGATGCGTTCCGTGTATCTTTTATAGCTGGCCCCCTTATATCTAACACAGTTTGTGTG	7800
Dd	7740	GGATGCGTTCCGTGTATCTTTTATAGCTGGCCCCCTTATATCTAACACAGTTTGTGTG	7799
OY	7801	GGCCATGTGTCACTGCTATCTATCTTATCATCTCTTATGGCTTAATAGTGTCTTC	7860
Dd	7800	GGCCATGTGTCACTGCTATCTATCTTATCATCTATCTTATGGCTTAATAGTGTCTTC	7859
OY	7861	TGTGTGATTAACAACCACTTTCTGTTCATTTACTGATGAATTTGTGCCCCACCCCCAC	7920
Dd	7860	TGTGTGATTAACAACCACTTTCTGTTCATTTACTGATGAATTTGTGCCCCACCCCCAC	7919
OY	7921	CTTTTTTTTTTTTTATTTTGAGACAAGGCTTTCTGCTGTATATCTTGGCAATCTTGGCTGCTC	7980
Dd	7920	CTTTTTTTTTTTTTATTTTGAGACAAGGCTTTCTGCTGTATATCTTGGCAATCTTGGCTGCTC	7979
OY	7981	GAGCTCACTCTGTAGACACAGGCTGTGAGGCTGTCTCCACTTTTACACTCCTGTGAAC	8040
Dd	7980	GAGCTCACTCTGTAGACACAGGCTGTGAGGCTGTCTCCACTTTTACACTCCTGTGAAC	8039
OY	8041	AGAGTAGCCATTAACCTTCAAGACAATTTTCTGTGTTTGGTTTGTATTAACATTTCTGTG	8100
Dd	8040	AGAGTAGCCATTAACCTTCAAGACAATTTTCTGTGTTTGGTTTGTATTAACATTTCTGTG	8098

OY	8101	TGATAGGCTGTATATGTGCAATGTTGTGTCTACAGTGTCTACAGTATGTGTATCTGTGTG	8158
Dd	8099	TGATAGGCTGTATATGTGCAATGTTTGTGTCTACAGTGTCTACAGTATGTGTATCTGTGTG	8158
OY	8161	TGGGACAGAGACAACAACCGAATGTGCCATTTCTCTCGATACACTACGATCTTGTAAATATGA	8220
Dd	8159	TGGGACAGAGACAACAACCGAATGTGCCATTTCTCTCGATACACTACGATCTTGTAAATATGA	8218
OY	8221	TGTATTATGTATGTATTATTTATAGTGTGCCAAGTATGACAGGTATTTTGTGGAGTTTTCAC	8280
Dd	8219	TGTATTATGTATGTATTATTTATAGTGTGCCAAGTATGACAGGTATTTTGTGGAGTTTTCAC	8278
OY	8281	CTTTCCTTGTGGGCTCTCCGCAATTAACCTACGCTCTCGGGCTATGTAGCAATGCCCTTCA	8340
Dd	8279	CTTTCCTTGTGGGCTCTCCGCAATTAACCTACGCTCTCGGGCTATGTAGCAATGCCCTTCA	8338
OY	8341	CTCGATGAGGCATCTGCTGTGCCCTGCTGTGCACCTCTCTCTTATTTTCCAGATGGGACTA	8400
Dd	8339	CTCGATGAGGCATCTGCTGTGCCCTGCTGTGCACCTCTCTCTTATTTTCCAGATGGGACTA	8398
OY	8401	CGCACTGCACTGTGCGCTTAAAGCTCACCAAGTCATCCAGATGTGGCTAGGCGAGGACACTCAG	8460
Dd	8399	CGCACTGCACTGTGCGCTTAAAGCTCACCAAGTCATCCAGATGTGGCTAGGCGAGGACACTCAG	8458
OY	8461	GGATATGTGTGGCTCTGTGCTTCCACAGTGTCTAGAAATTAACAGGSCATATCATCTGCTGAAG	8520
Dd	8459	GGATATGTGTGGCTCTGTGCTTCCACAGTGTCTAGAAATTAACAGGSCATATCATCTGCTGAAG	8518
OY	8521	ATTTTAACTGAAATCTCGAGGATAGAGCGAGGCACTCTACCAATGAGGGGTCTTTTGT	8580
Dd	8519	ATTTTAACTGAAATCTCGAGGATAGAGCGAGGCACTCTACCAATGAGGGGTCTTTTGT	8578
OY	8581	GTTTGGTTTGGTTTCCCTCTGCATTAAGATCAGGCAAGCTGTGAATATGTATAGCTTGGGCTAC	8640
Dd	8579	GTTTGGTTTGGTTTCCCTCTGCATTAAGATCAGGCAAGCTGTGAATATGTATAGCTTGGGCTAC	8638
OY	8641	ATTAACATCTTGTCTCAAAAAGCCTATATAGAGTATAGGAGGTGTGAGGCTTAAAGAAAGCCTT	8700
Dd	8639	ATTAACATCTTGTCTCAAAAAGCCTATATAGAGTATAGGAGGTGTGAGGCTTAAAGAAAGCCTT	8698
OY	8701	AAGCCGGCTGTGATAGCACACAGGATAGCCTGTGACATTAATAGCAAGACTTGTGTTCAAAA	8760
Dd	8699	AAGCCGGCTGTGATAGCACACAGGATAGCCTGTGACATTAATAGCAAGACTTGTGTTCAAAA	8758
OY	8761	ACATGAGAGGAGGGGTATGTTTTATGTGTGGGCTGTGTAAACAGGCACTAAAGGAGCCAA	8820
Dd	8759	ACATGAGAGGAGGGGTATGTTTTAAAGTGTGGGCTGTGTAAACAGGCACTAAAGGAGCCAA	8818
OY	8821	TGTAGACATTTGACTTAAGAAAGATCATATCAAAAGCCGGGTGGGCGAGGATAGAGGTTGG	8880
Dd	8819	TGTAGACATTTGACTTAAGAAAGATCATATCAAAAGCCGGGTGGGCGAGGATAGAGGTTGG	8878
OY	8881	ACTACAGTGTGTCAAGACCCCATATAGGAAGCAGTTTCCCTTCTTCCCTGTGGGCTCAAGC	8940
Dd	8879	ACTACAGTGTGTCAAGACCCCATATAGGAAGCAGTTTCCCTTCTTCCCTGTGGGCTCAAGC	8938

QY	8101	TGTAAGGTGATATATGTCATGTTTGGTTCCTCAGGTGCTCATCATGTGTAACCGTGTG	8160
Db	8099	TGTAAGGTGATATATGTCATGTTTGGTTCCTCAGGTGCTCATCATGTGTAACCGTGTG	8158
QY	8161	TGGACACAGAAACAACCCGATGTGCCATTCCTCAGATTAACGATCTTGTGTAATATGTA	8220
Db	8159	TGGACACAGAAACAACCCGATGTGCCATTCCTCAGATTAACGATCTTGTGTAATATGTA	8218
QY	8221	TGTATATATGTAATGTTTATATTTAGTGTGCCCAAGTATGCAGGTATATTTTGTGGAGTTTTCAC	8280
Db	8219	TGTATATATGTAATGTTTATATTTAGTGTGCCCAAGTATGCAGGTATATTTTGTGGAGTTTTCAC	8278
QY	8281	CTTCCCTTGTGGGCTCTCCCGCATTAACATCAGTCTCTCGGGCTACTGAGACATGCTCTTCA	8340
Db	8279	CTTCCCTTGTGGGCTCTCCCGCATTAACATCAGTCTCTCGGGCTACTGAGACATGCTCTTCA	8338
QY	8341	CTCGATGAGCCATCTCGCTCCCTCGCTGCGACCTCTCTTATTTCCAGATGGGACTA	8400
Db	8339	CTCGATGAGCCATCTCGCTCCCTCGCTGCGACCTCTCTTATTTCCAGATGGGACTA	8398
QY	8401	CGCAGCTGACATGCGCTTAAAGCTCAGCAAGTCATCCAGATGGCTAGCAGGAGACCTCAG	8460
Db	8399	CGCAGCTGACATGCGCTTAAAGCTCAGCAAGTCATCCAGATGGCTAGCAGGAGAGCTCAG	8458
QY	8461	GGATATGCTGGCTCTCTGCTCCACAGTGTAGATTTACAGGCTATCATCTACTGCTGGAG	8520
Db	8459	GGATATGCTGGCTCTCTGCTCCACAGTGTAGATTTACAGGCTATCATCTACTGCTGGAG	8518
QY	8521	ATTTTAAACCGAATCCTGAGGATAGGAGCAGGACCTCTCCAAATGAGGAGGTTCTTTTGT	8580
Db	8519	ATTTTAAACCGAATCCTGAGGATAGGAGCAGGACCTCTCCAAATGAGGAGGTTCTTTTGT	8578
QY	8581	GTTTGGTTGGTTCCTCTCTGCAATAGATCAGGCAAGTCTGAAATAGTGTAGGCTGGGCTAC	8640
Db	8579	GTTTGGTTGGTTCCTCTCTGCAATAGATCAGGCAAGTCTGAAATAGTGTAGGCTGGGCTAC	8638
QY	8641	ATTAACATCTTGTCTCAAAAAGCCTATAGAGGTAAGGAGGTGCAGGCTTAAGAGAGCCTT	8700
Db	8639	ATTAACATCTTGTCTCAAAAAGCCTATAGAGGTAAGGAGGTGCAGGCTTAAGAGAGCCTT	8698
QY	8701	AAGCGCGGTGATAGACACAGAGATAGCCTGACATATATTACAGACCTTGTTTCAAA	8760
Db	8699	AAGCGCGGTGATAGACACACAGAGATAGCCTGACATATATTACAGACCTTGTTTCAAA	8758
QY	8761	ACATGAGGAGGAGGGTATGTTTTAAGTGTGGCTGTGTAAACAGGCACTAAGGGAGCCAA	8820
Db	8759	ACATGAGGAGGAGGGTATGTTTTAAGTGTGGCTGTGTAAACAGGCACTAAGGGAGCCAA	8818
QY	8821	TGTAAGCATTTTGACTTAAGAAAGATATCATCAAAAGCCGGGTGAGAGGTTGG	8880
Db	8819	TGTAAGCATTTTGACTTAAGAAAGATATCATCAAAAGCCGGGTGAGAGGTTGG	8878
QY	8881	ACTACAGAGGTCAAAGACCCCATAGGAAGCAAGTTCCCTTCTCTGGAGCCCTCAAGC	8940
Db	8879	ACTACAGAGGTCAAAGACCCCATAGGAAGCAAGTTCCCTTCTCTGGAGCCCTCAAGC	8938
QY	8941	CTGGCTCGAGGCGCACTGCTCTCATATGCTTCTCCTTAGGCTCGTCACCATG	8995
Db	8939	CTGGCTCGAGGCGCACTGCTCTCATATGCTTCTCCTTAGGCTCGTCACCATG	8993
RESULT 3			
AAAF25499	ID	AAF25499 standard; DNA; 23024 BP.	
XX	AAAF25499;		
AC			
XX			
XX	15-MAY-2001 (first entry)		
DE			
XX	Nucleotide sequence of a murine ABCA7 (ABCA7) transporter.		
XX			
XX	ABCA transporter; high-density lipoprotein cholesterol;		
KM	ABCA7 transporter; ss.		

XX SS Mus musculus.  
XX PN WO200109314-A1.  
XX PD 08-FEB-2001.  
XX PE 28-JAN-2000; 2000WO-FR00209.  
XX PR 30-JUL-1999; 99FR-0009926.  
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX PI Chimini G;  
XX DR WPI; 2001-182953/18.  
XX PT Selecting agents that modulate ABCA transporters, useful e.g. for  
PT normalizing serum cholesterol levels, comprises using transgenic  
PT animals with an inactive ABCA gene allele -  
XX PS  
PS Claim 12; Fig 22A; 113pp; French.  
XX CC The specification describes a method for selecting or screening agents  
CC that modulate ABCA transporters. The method comprises using non-human  
CC recombinant mammals with an inactivated allele of the gene encoding the  
CC ABCA transporter, or cells with an inactivated allele of ABCA  
CC transporter, from any tissue of the recombinant mammal, preferably with  
CC an allele truncated in one or both exons corresponding to the first  
CC and/or second ATP-binding cassettes (NBD or 2). Compounds that stimulate  
CC ABCA transporters may be useful for increasing (normalizing) serum  
CC levels of high-density lipoprotein cholesterol. The present sequence  
CC represents the genomic sequence of a murine ABCA7 (ABCA7) transporter.  
XX SO  
Sequence 23024 BP; 5195 A; 6205 C; 6348 G; 5266 T; 10 other:  
SO  
Query Match 1.7%; Score 153.2; DB 22; Length 23024;  
Best Local Similarity 79.1%; Pred. NO. 1.7e-26;  
Matches 182; Conservative 0; Mismatches 48; Indels 0; Gaps 0  
QY 4552 ATAGGCGAGACTTTGTCTTAAGAAAAAATGAAGAGCCGACGATGTCGACACAGCCT 4611  
DB 16839 ACACGAGCAAAACCCGTCTAAAAAACCAGAAACAGGCGGCGAGTGGTGTCAGCCT 16898  
QY 4612 TTAATCCGACCACTTGGAGGAGCAGAGCAGAGATTTCTGATTCAGGTCAGCAGCCTGCT 4671  
DB 16899 TTAATCCGACCACTTGGAGGAGCAGAGCAGAGTGATTTCTGATTCGAGGCTAAGCCTGGT 16958  
QY 4672 CTATAGAGTGAATTCCTCAGAGACAGCCAGGCGTACACAGAGAAACCCCTGTTTGAANAACCA 4731  
DB 16959 CTACATAGTGAATTCCTCAGAGACAGCCAGAGCTACACAAAGAAACCCGTGTTTGAANAACAA 17018  
QY 4732 GAAAAACAAACAAACAAACAAACAAACAAACCCCAACCCCAACCAAC 4781  
DB 17019 ACAAACAAACAAACCAACCTAAAAAACCAGAAATAGTAACAAACAAACCAAC 17068  
RESULT 4  
ID AAL38337 standard; DNA: 215980 BP.  
XX AAL38337;  
XX AC  
XX 15-AUG-2002 (first entry)  
XX DT  
XX Complementary strand of a genomic sequence encoding a mouse Ngr3.  
DE Cerebroprotective; neuroprotective; cytosstatic; Nogo receptor homologue;  
KM Ngr3; Ngr3; axonal growth; central nervous system; CNS; cerebral injury;  
KM spinal cord injury; stroke; demyelinating disease; multiple sclerosis;  
KM multiphasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;  
KM multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;  
KM

XX	OS	Mus sp.	Alexander's disease; Canavan's disease; metachromatic leukodystrophy;
XX	XX		Krabbe's disease; immune; bait protein; genetic mapping; gene therapy;
XX	KM	transgenic animal; unregulated cellular growth; cancer; tumour; mouse;	
XX	XX	murine; ds.	
XX	PD	11-APR-2002.	
XX	PF	06-OCT-2001; 2001MO-US31488.	
XX	PR	06-OCT-2000; 2000US-238361P.	
XX	XX		
XX	PA	(UYU ) UNIV YALE.	
XX	XX	(BIO ) BIOGEN INC.	
XX	PI	Stiltmacter SM, Cate RL, Sah DMW;	
XX	XX	WPI; 2002-416677/44.	
XX	PT	Novel Nogo receptor homolog polypeptide, Ngr2 or Ngr3, useful for	
XX	PT	treating central nervous system disorder, cerebral injury, spinal cord	
XX	PI	injury, stroke, and demyelinating diseases	
XX	XX		
XX	XX	Example 16; Page 219-275; 277pp; English.	
XX	CC	The invention relates to a Nogo receptor homologue polypeptide, Ngr2 or	
XX	CC	Ngr3, comprising a 50 amino acid LRRT sequence, a 264 amino acid NLRRT	
XX	CC	sequence, or a 420, 461 or 392 amino acid sequence, all given in the	
XX	CC	specification. The Ngr3 protein or its binding antibody is useful for	
XX	CC	decreasing inhibition of axonal growth of a central nervous system (CNS)	
XX	CC	neuron, by contacting the neuron Ngr3 or its antibody, and for treating	
XX	CC	CNS disease, disorder or injury. Ngr3 or a vector comprising Ngr3 is	
XX	CC	useful for treating cerebral injury, spinal cord injury, stroke,	
XX	CC	demyelinating diseases, e.g. multiple sclerosis, monoplasic	
XX	CC	demyelination, encephalomyelitis, multifocal leukoencephalopathy,	
XX	CC	panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,	
XX	CC	Alexander's disease, Canavan's disease, metachromatic leukodystrophy and	
XX	CC	Krabbe's disease. Ngr3 is useful for inducing an immune response in a	
XX	CC	mammal against Ngr3, as a bait protein in a two-hybrid or three-hybrid	
XX	CC	assay, and as a research tool for identification, characterisation and	
XX	CC	purification of interacting, regulatory proteins. The nucleotide	
XX	CC	sequences of the invention are useful for screening for RFLP associated	
XX	CC	with certain disorders, for genetic mapping, and for gene therapy. The	
XX	CC	vector containing Ngr3 is useful for producing non-human transgenic	
XX	CC	animals. The Ngr3 binding antibody is useful for isolating and purifying	
XX	CC	Ngr3, for localisation and/or quantitation of Ngr3, and for diagnostic	
XX	CC	and therapeutic purposes. The sequences of the invention, vectors and	
XX	CC	antibodies are useful for treating or preventing unregulated cellular	
XX	CC	growth such as cancer and tumour growth. This polynucleotide sequence	
XX	CC	represents the	
XX	XX		
XX	SQ	Sequence 215980 BP; 56872 A; 50995 C; 50240 G; 54437 T; 3436 other;	
XX	XX		
XX	XX	Query Match 1.7%; Score 152.2; DB 24; Length 215980;	
XX	XX	Best Local Similarity 82.2%; Pred. No. 9,6e-26;	
XX	XX	Matches 175; Conservative 0; Mismatches 38; Indels 0; Gaps 0;	
XX	QY	4560 AGACTTTGTCCTTAAGAAAAAATGAAAGCCAGCAGAGTGATGGACACGCGTTTAATGCC 4619	
XX	DB	119710 ATATATTTTACTACATTTAAAAAAACAAAGCGCGGTGTGGCGCATCTTAATGCC 119769	
XX	QY	4620 AGCACTTGGGAGGACAGAGCAGAGATTTTGTGATTTCAAGGCCAGCGCTGTATAGG 4679	
XX	DB	119770 AGCACTTCGGGAGGACAGAGCAGCGGATTTGTGATTTCAAGGCCAGCGCTGTATCAAG 119829	
XX	QY	4680 TAGATTTCCAGAGCAGAGCGGTACACAGAGAAACCTTTTGAAGAAACAGAAANACA 4739	
XX	DB	119830 TGAGTTCCAGGACAGCGGCTACACAGAGAAACCTTCTTCGAAAAACCAAAAAA 119889	
XX	QY	4740 AATCAAAACAAAAACAAACCAACCA 4772	



Db 119890 AAAAAAAAAAACCACCAACACACA 119922

RESULT 5  
AAC82696

XX  
AC AAC82696;

XX	5'	3'
DE	Murine variant Zalpha32 cDNA	

KM Zalpha32, murine; secreted alpha-helical protein-32; antinflammatory  
 KW spermatogenesis; steroidogenesis; testicular differentiation;  
 KW regulatory control; hypochalamic- pituitary-gonadal axis;  
 KW reproductive system disorder; immunological system disorder; ss.

OS Mus musculus.

	Key	Location/Qualifiers
PH		19..624
FT	CDS	
FT		/*tag= a
FT		/product= "variant Zalpha322

PN WO200071717-A1

PD 30-NOV-2000.

PF 25-MAY-2000; 2000WO-US14563.

PR 26-MAY-1999; 99US-0320159.

PA (ZYMO) ZYMOGENETICS INC.

PI Conk1 in DC, Gao Z;

DR WPI; 2001-032038/04.

DR P-PSDB; AAB45760.

PT Novel mammalian secreted alpha-helical protein-32 (Zalpha32) useful for  
 PT identifying antagonists of its activity that are used in treating  
 PT Zalpha32-induced inflammation -  
 XX  
 XX Disclosure: Page 62-63; 73pp; English.

CC This invention describes a novel mammalian secreted alpha-helical  
CC protein-32 (Zalpa32) (I). The invention also describes (1) an isolated  
CC polynucleotide (II) encoding (1); (2) an antibody (III) that specifically  
CC binds to (1); and (3) an educational kit (IV) for the teaching of  
CC molecular biology and/or biochemistry comprising (II). The products of  
CC the invention have antiinflammatory activity. (III) as an antagonist to  
CC zalpa32 is useful for treating zalpa32-induced inflammation. Zalpa32  
CC can be used to identify inhibitors of its activity and to prepare  
CC antibodies that specifically binds to zalpa32 epitopes. Antibodies or  
CC polypeptides, directly or indirectly conjugated to drugs or toxins are  
CC useful for in vivo diagnostic or therapeutic applications.  
CC Zalpa32-cytokine fusion proteins are useful for enhancing in vivo  
CC killing of target tissue. (I) and (II) are useful to identify and isolate  
CC receptors involved in spermatogenesis, steroidogenesis, testicular  
CC differentiation and regulatory control of the  
CC hypothalamic-pituitary-gonadal axis and also for testing disorders of the  
CC reproductive system and immunological systems.

Sequence 1445 BP; 418 A; 319 C; 398 G; 310 T; 0 other;

Query Match	1.7%	Score 150.6	DB 22	Length 1445
Best Local Similarly	83.4%	Pred. No. 1.8e-26		
Matches 171	Conservative	0	Mismatches 34	Indels 0
				Gaps 0

[illegible]

Db	988	TTTAAAAAATGTGTGCTTACCGCGGCGTGTGGCCGACGCCCTTTATCCACACTTGGG	1047
QY	4630	AGGCAGAGCAGCGCAGATTTTCTAGATTCAAGGCGACGCTGTCTATAGTGTAGTTCCAG	4689
Db	1048	AGGCAGAGCGCGCGGATTTTCTGAGTTTGGAGGCGACGCTGTCTAGAGTGTAGTTCCAG	1107
QY	4650	GACGCGCAGGGCTTACACAGAGAAACCCCTGTTTGGAAAAACCGAAACCAAAACA	4749
Db	1108	GACGCGCAGGGCTTACACAGAGAAACCCCTGTTTGGAAAAACCAAAACCAAAACA	1167
QY	4750	AAACAAACCAAAACCCCAACCCAAA	4774
Db	1168	ACAAACAAAAAAACCAAAACCAAAA	1192

RESULT 6  
AAA40866  
ID AAA40866 standard; DNA; 7208 BP

AC AAAA40866;

DT 16-AUG-2000 (first entry)

DE Murine tumour necrosis factor alpha (TNFalpha) nucleotide sequence

KM Antisense oligonucleotide: phosphorothioate; TNFalpha; cytokine; inhibitor  
KM tumour necrosis factor alpha: inflammatory bowel disease; diabetes;  
KM rheumatoid arthritis; infectious disease; multiple sclerosis; hepatitis;  
KM paracitavilis; atopic dermatitis; allograft rejection;  
KM autoimmune disease; inflammatory disease; ds.  
KM

OS Mus sp.

PN WO2000020645-A1

PD 13-APR-2000

PF 05-OCT-1999; 99WO-US23205.

PR 05-OCT-1998; 98US-0166186.

PR 18-MAY-1999; 99US-0313932.

PA (ISIS-) ISIS PHARM INC.

PI Baker BF, Bennett CF, Butler MM, Shanahan WJ;

DR WPI; 2000-303808/26.

DR P-P\$DB; AAY88596.

PT Oligonucleotide for treating diseases associated with human tumour  
PT necrosis factor- $\alpha$  (TNF $\alpha$ ) such as, diabetes and rheumatoid  
PT arthritis, comprises nucleotide sequence complementary to intron 5'  
PT nucleic acid encoding TNF $\alpha$

PS Example 8; Page 173-179; 283pp; English

CC This sentence represents the murine tumour necrosis factor alpha  
CC (TNFalpha) nucleotide sequence. TNFalpha is an important cytokine that  
CC plays a role in host defence. It is produced mainly in macrophages and  
CC monocytes in response to infection, invasion, injury or inflammation.  
CC Overexpression of TNFalpha can result in disease states, particularly in  
CC infectious, inflammatory and autoimmune diseases. The invention relates  
CC to antisense oligonucleotides which are capable of modulating the  
CC TNFalpha gene expression. The oligonucleotides optionally have a  
CC phosphorothioate backbone, and may also optionally contain at least one  
CC 2'-O-methoxyethyl modification. The oligonucleotides are useful for  
CC modulating the expression of human TNFalpha in cells and tissues,  
CC reducing a human cell inflammatory response, reducing the blood glucose  
CC level in a human and treating a human having a disease or condition  
CC associated with TNFalpha. Examples of diseases associated with TNFalpha  
CC include diabetes, inflammatory bowel disease, multiple sclerosis,  
CC paracellitis, rheumatoid arthritis, infectious disease, hepatitis, atopic  
CC dermatitis or allograft rejection. The antisense oligonucleotides are



CC also useful for modulating the function of a selected nucleic acid  
CC sequence in adipose tissue.  
XX  
SQ Sequence 7208 BP; 1750 A; 2011 C; 1810 G; 1637 T; 0 other;  
Query Match 1.7%; Score 148.8; DB 21; Length 7208;  
Best Local Similarity 88.0%; Pred. No. 1.1e-25;  
Matches 162; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 4574 AGAAAAAATGAACCCAGCAGTGTGGCAGACCCCTTTAATCCAGACCTGGAGGC 4633  
D 429 AAAAAAAAAAAAAAGCTGGCGTGTGGCACACCTTTAATCCAGACCTGGAGGC 488  
QY 4634 AGAAGCAGGAGATTTCTGAGTTCAAGCCAGCCCTGCTATAGAGTGCAGAGCA 4693  
D 489 AGAGCAGGCGGATTTCTGAGTTCAAGCCAGCCCTGCTATAGAGTGCAGAGCA 548  
QY 4694 GCCAGGCTACACAGAGAAACCCCTTTTGAACCAAGAAAAACAAACAAAC 4753  
D 549 GCCAGGCTACACAGAGAAACCCCTGCTCGAAGAAAAAGCAAAAAA 608  
QY 4754 AAAA 4757  
D 609 AAAA 612  
RESULT 7  
AAF63436/C  
ID AAF63436 standard; DNA; 6789 BP.  
XX  
AC AAF63436;  
XX  
DT 14-MAY-2001 (first entry)  
XX  
DE Murine CD39-L4 genomic DNA sequence.  
XX  
KW Human CD39-like protein; apyrase; NDase; platelet function inhibitor;  
KW myocardial infarction; cerebral ischemia; angina; arterial thrombosis;  
KW cerebral artery thrombosis; platelet aggregation; inflammation;  
KW apoptosis; autoimmune disorder; neurological disorder;  
KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2; ds.  
XX  
OS Mus sp.  
XX  
PN MO200110205-A1.  
XX  
PD 15-FEB-2001.  
XX  
PF 09-AUG-2000; 2000MO-US21790.  
XX  
PR 09-AUG-1999; 990US-0370265.  
PR 11-JAN-2000; 2000US-0481238.  
PR 25-APR-2000; 2000US-0557800.  
PR 26-MAY-2000; 2000US-0583231.  
PR 30-JUN-2000; 2000US-0608285.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
P1 Ford J, Mulero JJ, Yeung G;  
XX  
DR WPI; 2001-147489/15.  
XX  
PT Polynucleotides encoding human CD39-like polypeptides, with apyrase  
PT and/or NDase activity, which are useful in the treatment of  
PT pathological conditions caused by thrombosis (e.g. myocardial  
PT infarction) and inflammatory disorders -  
XX  
PS Example 21; Page 108-110; 203pp; English.  
XX  
CC This invention relates to polynucleotides encoding human CD39-like  
CC polypeptides with apyrase and/or NDase activity. The polypeptides having  
CC ATPase, including NDase, activity are useful for inhibiting platelet  
CC function and can therefore be used in the prophylaxis or treatment of

CC pathological conditions caused by or involving thrombosis or excessive  
CC coagulation or excessive platelet aggregation, such as myocardial  
CC infarction, cerebral ischemia, angina, arterial thrombosis, cerebral  
CC artery thrombosis or intracardiac thrombosis, and conditions associated  
CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
CC modulating disease states (including platelet aggregation, inflammation  
CC and apoptosis) associated with ADP or other purinergic signaling by  
CC reducing the levels of NDps. The polypeptides are also useful for  
CC prophylaxis or treatment of inflammation related disorders, such as  
CC disorders involving sepsis or systemic inflammatory response syndrome or  
CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,  
CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;  
CC neurological disorders including neurodegenerative diseases, epilepsy,  
CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
CC sequence represents the murine CD39 like protein CD39-L4 genomic DNA  
CC sequence.  
XX  
SQ Sequence 6789 BP; 1775 A; 1476 C; 1582 G; 1952 T; 4 other;  
Query Match 1.6%; Score 145.6; DB 22; Length 6789;  
Best Local Similarity 62.3%; Pred. No. 6.8e-25;  
Matches 304; Conservative 0; Mismatches 164; Indels 20; Gaps 4;  
QY 1854 GGAGGAGATGGCTCGACTTCCAGAGCACCTTCTCTTGCAGAGGACTGATTC 1913  
D 1483 GGCCAGAGAGATGACTCGAAGATTAGAGTATGCCCTCTCTTACAGAGGACTGATTC 1424  
QY 1914 AGTTCCAGAGA--CTCATATGGTGGCTCACAGCCATCTGTAAATCCAGTCCAGAGGTT 1971  
D 1423 AGTTCCAGAGATGCTCAAGAGGTAGTTGTATAGCTCTATATATTCAGATCCAGAGCATC 1364  
QY 1972 CCACACCCCTTCTTGCCCTCCACAGGACACATATAGTACAGACATATCATGAGG 2031  
D 1363 TGACACCCCTCATGCAAGTACACCCACACACAAATTAATAATAATTCATATTTG 1304  
QY 2032 CAAACACCCATACACATATAATAATAAGAACTTAAAGTGATGTTGGTAAA 2091  
D 1303 AAGAAACCTAAAGTGTAGCATTTCAATTAATGAACCAAGTCTACCTGGTTGCTTT 1244  
QY 2092 CATGTGCTTACACATGCTGATTAAGACATGTACACAGCCACACTGAAAGGATCTG 2151  
D 1243 AAACATTTGTTTTTTTTTTTCTAGATPAGCACTAATAATAAACAATAATG-----G 1190  
QY 2152 GGGCTGAGAGATGGCTCAGCGGTTAAGAGCACTGCTCTTCCGAAGAGGTCCTG 2211  
D 1189 GGGCTGTCAAGATGGCTCAGAGGGTAAAGAGCACTGCTCTTCTTC-----GAAGTCTCTG 1134  
QY 2212 AGTTCAATTCCTAGCACACATGCTGCTCACAACATCCATTAATGAGATCTGACACC 2271  
D 1133 AGTTCAATTCCTAGCACACATGCTGCTCACAACATCCATTAATGAGATCTGACACC 1074  
QY 2272 TCTTCTGTGATCTGAGAGACAGTGCAGAGGTATACAGTCTTATGATTAATAAT 2331  
D 1073 TCTTGTGTGATCTAAAGTC-----AGCTACAGTGTACTTATGATTAATAAT 1022  
QY 2332 AAATCTTT 2339  
D 1021 AAATCTTT 1014  
RESULT 8  
AAA47451/C  
ID AAA47451 standard; CDNA; 2221 BP.  
XX  
AC AAA47451;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Mouse TANGO 232 coding sequence.  
XX



```
Oy 4562 ACTTGTCTTTAAGAAAAATGAAAGCCACAGTGGTGGCAGACGCTTTAATCCG 4621
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4668 ATTCTCTTTTAAAAAAGAAAAAGACCGCGCATGGTGGACATGCTTTAATCCAG 4727
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4622 CACTTGGAGCGCAGAGCAGAGCATTTCTGATCAAGGCCAGCCTGGTCTATAGAGTG 4681
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4728 CACTTGGAGCGCAGAGCAGAGCATTTCTGATCAAGGCCAGCCTGGACACAGAGTG 4787
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4662 AGTTCAGACAGCAGCGCTACACAGAGAAACCTGTTTGAATAACCAAGAAAAACA 4741
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4788 AGTTCAGACAGCAGCGCTACTCAGAGAACTGTCTTGAAAAAAGAGAGGA 4847
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4742 ACAAACAACACA 4754
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4848 GAAAGAAAAAAGA 4860
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
AA23318
ID AAX23318 standard; DNA: 1698 BP.
AC AAX23318;
XX
XX 11-JUN-1999 (first entry)
DE Mouse 1-alpha-OHase promoter region DNA #3.
XX
XX 1-alpha-OHase: vitamin D 1-alpha-hydroxylase; vitamin D pathway;
KW renal cytochrome P450 enzyme; 25-hydroxyvitamin D; human; mouse; rat;
KW diagnosis; treatment; disorder; endocrine; promoter; ss.
XX
XX Mus sp.
OS
XX WO9907835-A2.
PN
XX 18-FEB-1999.
PD
XX 06-AUG-1998; 98WO-CA00758.
PF
XX 06-AUG-1997; 97US-0906791.
PR
XX (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
PA
XX PI Glorieux FH, St-Arnaud R;
XX
XX WPI: 1999-190048/16.
DR
XX New polynucleotides encoding 25-hydroxyvitamin D 1- $\alpha$ -hydroxylase -
PT useful to detect and treat vitamin D-related disorders
PT
XX
XX Claim 12; Fig 15; 85pp; English.
PS
XX
XX This invention describes novel 25-hydroxyvitamin D 1-alpha-hydroxylase
CC (1-alpha-OHase), a renal cytochrome P450 enzyme of the vitamin D
CC pathway, and polynucleotides encoding the enzyme from human, mouse and
CC rat tissue. Molecules of the invention are used to diagnose and treat
CC vitamin D-related disorders, to produce vitamin D metabolites, and to
CC identify modulators of the enzyme expression. The invention describes
CC an animal cell transformed with the enzyme promoter which is used to
CC identify compounds which modulate activity of the polynucleotide
CC encoding the enzyme, and to modulate the production of vitamin D in
CC patients with vitamin D endocrine disorders in the vitamin D
CC hydroxylation pathway, the enzyme catalyses conversion of
CC 25-hydroxyvitamin D to 1 alpha 25(OH)2D3.
CC
XX
XX Sequence 1698 BP; 499 A; 425 C; 414 G; 360 T; 0 other;
SQ

Query Match 1.6%; Score 141.4; DB 20; Length 1698;
Best Local Similarity 77.1%; Pred. No. 3.6e-24;
Matches 172; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Oy 4552 ATAGAGCAGACTTGTCTTTAAGAAAAAATGAAGCCACAGTGGTGGCAGACGCT 4611
```

```
Db 923 AAAAGGCAATTTGTCATATTAACAGTGAATCAAGCGCGGGCGTGGTGGACAGGCT 982
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4612 TTAATCCAGCAGCTTGGAGCGCAGAGCAGCAGATTTTCTGATTCAGCCAGCCTGGT 4671
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 983 TTAATCCAGCAGCTTGGAGCGCAGAGCGCGCTGATTTCTGATTCATGCGCAGCCTGGT 1042
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4672 CTATGAGTGAAGTTCAGAGCAGCGCTACACAGAGAAACCTGTTTGAATAACCA 4731
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1043 CTACAGAGTGAAGTTCAGAGCAGCGCTTACACAGAGAAACCTGTCTGGGAAAAACA 1102
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4732 GAAAAACAACCAAAACAACCAAAACAACCAAAACCAAA 4774
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1103 AAAAAACAACCAAAACAACCAAAACCAAAACCAAGTACTTAA 1145
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
AA23316
ID AAX23316 standard; DNA: 4105 BP.
AC AAX23316;
XX
XX 11-JUN-1999 (first entry)
DE Mouse 1-alpha-OHase promoter region DNA #2.
XX
XX 1-alpha-OHase: vitamin D 1-alpha-hydroxylase; vitamin D pathway;
KW renal cytochrome P450 enzyme; 25-hydroxyvitamin D; human; mouse; rat;
KW diagnosis; treatment; disorder; endocrine; promoter; ss.
XX
XX Mus sp.
OS
XX WO9907835-A2.
PN
XX 18-FEB-1999.
PD
XX 06-AUG-1998; 98WO-CA00758.
PF
XX 06-AUG-1997; 97US-0906791.
PR
XX (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
PA
XX PI Glorieux FH, St-Arnaud R;
XX
XX WPI: 1999-190048/16.
DR
XX New polynucleotides encoding 25-hydroxyvitamin D 1- $\alpha$ -hydroxylase -
PT useful to detect and treat vitamin D-related disorders
PT
XX
XX Example 4; Page 79-81; 85pp; English.
PS
XX
XX This invention describes novel 25-hydroxyvitamin D 1-alpha-hydroxylase
CC (1-alpha-OHase), a renal cytochrome P450 enzyme of the vitamin D
CC pathway, and polynucleotides encoding the enzyme from human, mouse and
CC rat tissue. Molecules of the invention are used to diagnose and treat
CC vitamin D-related disorders, to produce vitamin D metabolites, and to
CC identify modulators of the enzyme expression. The invention describes
CC an animal cell transformed with the enzyme promoter which is used to
CC identify compounds which modulate activity of the polynucleotide
CC encoding the enzyme, and to modulate the production of vitamin D in
CC patients with vitamin D endocrine disorders in the vitamin D
CC hydroxylation pathway, the enzyme catalyses conversion of
CC 25-hydroxyvitamin D to 1 alpha 25(OH)2D3.
CC
XX
XX Sequence 4105 BP; 1063 A; 1032 C; 1063 G; 947 T; 0 other;
SQ

Query Match 1.6%; Score 141.4; DB 20; Length 4105;
Best Local Similarity 77.1%; Pred. No. 5.7e-24;
Matches 172; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Oy 4552 ATAGAGCAGACTTGTCTTTAAGAAAAAATGAAGCCACAGTGGTGGCAGACGCT 4611
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3335 AAAAGCAAAATTTGTCTATTAACAGTGAATCAAGCCCGGCGCTGGTGGCAGCGCT 3394
```

Qy	4612	TTTATTCCTCCAGACTCTTGGGAGGCGAAGACAGCGCAGATTTCTAGTTCACAGGCCAGCTGTGT	4671
Db	3395	TTTATTCCTCCAGACTCTTGGGAGGCGAAGGCGCGTGTGATTTCTAGTTCACAGGCCAGCTGTGT	3454
Qy	4672	CTATAGAGTGTAGTTCCTCCAGGACGCGGCGCTACACGACGAACCCCTGTGTTTGAAACCA	4751
Db	3455	CTACAGAGTGTAGTTCCTCCAGGACGCGGCTTACACGAGAACCCCTGTCTGGGAAACCA	3514
Qy	4752	GAAACACAAACAAACAAACAAACAAACAAACCAACCCAA	4774
Db	3515	AAAAACAAACAAACAAACAAACAAACCCACAGTGTCTAA	3557

```

RESULT 12
ID AAL47852 standard; DNA: 20775 BP.
XX
XX AAL47852:
AC
CT 19-SEP-2002 (first entry)
XX
DE Murine basic helix-loop-helix BHLH-PAS protein related gene #2.
XX
KW Basic helix-loop-helix protein; BHLH-PAS protein; neuroprotective;
KW neotrophic; Per-Arnt-Sim homology domain; developmental disease;
KW cell proliferative disease; nerve function; dementia; gene; ds;
KW Alzheimer's disease; memory.
XX
OS Mus musculus.
XX
PN MO200253729-Al.
XX
PD 11-JUL-2002.
XX
PF 17-DEC-2001; 2001MO-JP11064.
PR 27-DEC-2000; 2000JP-0398548.
PR 19-MAR-2001; 2001JP-0077740.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PT Ohe N:
XX
DR WPI: 2002-528860/56.
XX
PT Mammalian basic helix-loop-helix bHLH-PAS proteins and encoding genes
PT with transcriptional regulatory ability, useful in diagnosis and
PT developing drugs for e.g. genetic diseases related to nerve functions,
PT motion and long-term memory.
XX
PS Example 10: Page 157-173; 176pp; Japanese.
XX
CC The present invention provides the protein and coding sequences of basic
CC helix-loop-helix Per-Arnt-Sim domain containing proteins (BHLH-PAS
CC proteins) from the human, mouse and rat. The sequences are useful in the
CC diagnosis and treatment of diseases associated with cell proliferation
CC and differentiation, development, functional expression, particularly
CC genetic diseases related to nerve function, motion and long-term memory
CC like dementia or Alzheimer's disease. The present sequence is a coding
CC sequence of the invention.
XX
SQ Sequence 20775 BP; 5384 A; 5246 C; 5100 G; 5045 T; 0 other;
XX
Query Match 1.6%; Score 140.8; DB 24; Length 20775;
Best Local Similarity 83.3%; Pred. No. 1.8e-23;
Matches 160; Conservative 0; Mismatches 32; Indels 0; Gaps 0
QY 4572 TAAGAAAAAATGAAGCCAGCAGTGTGGCACACGCGCTTTAAATCCAGCACTTGAGAG 4631
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19538 TAGAAATGATAGATTATGCCGGGGCGTGTGTGTGTCGCTTTATTCACGACTGCGGAG 19597
QY 4632 GCAGAACGACGACGATTTGTGAGTTCAAGGCCACGCGTGTCTTAAGATGAGTTCCAGGA 4691

```

Accession	Sequence	Position
Db 19598	GCAGAGGCAGAGGAGATTTCGTGAGTTCGAGCCAGCCTGGTCTACCAAGAGATTCACAGA	19657
Qy 4692	CAGCCAGGGCTACACAGAGAAACCTGTTTGAAGAACCGAAGAAACAAACAAACAAA	4751
Db 19658	CAGCCAGGGCTACACAGAGAAACCTGTCTCGAAGACCAACAAACAAACAAACAAA	19717
Qy 4752	ACAAAACCAAAAC 4763	
Db 19718	ACAAAACCAAAAC 19729	

ABN83212	standard; DNA; 20775 BP.
XX	ABN83212;
XX	13-SEP-2002 (first entry)
XX	Mouse transcriptional regulator gene #2.
XX	Mouse; transcriptional regulator; neurotrophic; neuroprotective; aging;
KW	cognition failure; dementia; Alzheimer's disease; gene; ds.
XX	Mus musculus.
OS	
XX	Key
FH	Location/Qualifiers
FH	9594..14342
CD	/tag= a
FT	/product= "transcriptional regulator"
FT	/note= "Contains 7 introns"
FT	9594..9768
FT	/tag= b
FT	/number= 1
FT	9769..10522
FT	/tag= c
FT	/number= 1
FT	10523..10674
FT	/tag= d
FT	/number= 2
FT	10675..10848
FT	/tag= e
FT	/number= 2
FT	10849..10951
FT	/tag= f
FT	/number= 3
FT	10952..11061
FT	/tag= g
FT	/number= 3
FT	11062..11335
FT	/tag= h
FT	/number= 4
FT	11336..11522
FT	/tag= i
FT	/number= 4
FT	11523..11632
FT	/tag= j
FT	/number= 5
FT	11633..11959
FT	/tag= k
FT	/number= 5
FT	11960..12095
FT	/tag= l
FT	/number= 6
FT	12096..12225
FT	/tag= m
FT	/number= 6
FT	12226..13655
FT	/tag= n
FT	/number= 7
FT	13656..14313
FT	/tag= o

```

FT      exon                               /number= 7
FT      /tag= 14314..14342
FT      /number= 8
FT      /number= 8
XX      WO200253736-A1.
XX      11-JUL-2002.
XX      17-DEC-2001; 2001WO-JP11063.
XX      27-DEC-2000; 2000JP-0398548.
XX      19-MAR-2001; 2001JP-0077740.
XX      (SUMO ) SUMITOMO CHEM CO LTD.
XX      Ohe N;
XX      WPI: 2002-528862/56.
XX      P-PSDB; ABB81801.
XX      Examining the ability to control nerve plasticity depending on a
XX      transcriptional regulator, useful in diagnosis and developing drugs for
XX      cognition failure due to aging, dementia and Alzheimer's disease
XX      Disclosure: Page 102-119; 127pp; Japanese.
XX      The invention relates to a novel method for examining the ability to
XX      control nerve cell plasticity depending on a transcriptional regulator.
XX      The method of the invention has nootropic and neuroprotective activity.
XX      The method is useful in diagnosis and developing drugs for improving or
XX      treating cognition failure due to aging, dementia and Alzheimer's
XX      disease. The sequence encodes the mouse transcriptional regulator of
XX      the invention.
XX      Sequence 20775 BP; 5384 A; 5246 C; 5100 G; 5045 T; 0 other;
XX      Query Match 1.6%; Score 140.8; DB 24; Length 20775;
XX      Best Local Similarity 83.3%; Pred. No. 1.8e-23;
XX      Matches 160; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
XX      4572 TAAGAAAAAATGAAGGCCACAGCAGTGGTGGCAGCAGCCTTTAAATCCAGCAGCTTGGAG 4631
XX      || || || || || || || || || || || || || || || || || || || || || || ||
XX      Db 19538 TAGAATGTGAAGATTATGCCGGCGTGGTGGTCTGCTTAATCCAGCAGCTGGGAG 19537
XX      4632 GCAGAGCAGCAGGATTTCTGAGTTCAAGGCCAGCCTGGTCTATAGAGTGAAGTCCAGGA 4691
XX      || || || || || || || || || || || || || || || || || || || || || || ||
XX      Db 19598 GCAGAGCAGCAGGATTTCTGAGTTCAAGGCCAGCCTGGTCTAACAAGTGAAGTCCAGGA 19657
XX      4692 CAGCCAGGAGCTACACAGAGAAACCTGTTTGAAGAAACAGAGAAACCAAAACAAACAA 4751
XX      || || || || || || || || || || || || || || || || || || || || || || ||
XX      Db 19658 CAGCCAGGAGCTACACAGAGAAACCTGTTTGAAGAAACCAAAACCAAAACAAACAA 19717
XX      4752 ACAAAACAAAC 4763
XX      || || || || || || || || || || || || || || || || || || || || || || ||
XX      Db 19718 ACAAAACAAAC 19729
XX      RESULT 14
XX      AAV19608/c
XX      ID AAV19608 standard; DNA; 3360 BP.
XX      AAV19608;
XX      11-SEP-1998 (first entry)
XX      Mus musculus EPCR gene promoter.
XX      EPCR; endothelial cell protein C binding protein receptor;
XX      gene therapy; atherosclerosis; vascular disease; drug screening;
XX      autoimmune disease; ss.
XX      Mus musculus.

```

```

XX      Key Location/Qualifiers
XX      misc_feature 2270..2840
XX      /tag= a
XX      /note= "large vessel endothelial response element"
XX      misc_feature 2990..3061
XX      /tag= b
XX      /note= "serum response element"
XX      misc_feature 3007..3014
XX      /tag= c
XX      /note= "thrombin response element"
XX      /tag= d
XX      /note= "endothelial specific element"
XX      WO9820041-A1.
XX      14-MAY-1998.
XX      07-NOV-1997; 97WO-US20364.
XX      04-AUG-1997; 97US-0054533.
XX      08-NOV-1996; 96US-0030718.
XX      (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX      Esmom CT, Gu J;
XX      WPI: 1998-286871/25.
XX      Regulatory elements from the endothelial protein C receptor promoter
XX      - useful to direct expression of genes or nucleotide molecules e.g.
XX      to endothelial cells or only large vessel endothelial cells in gene
XX      therapy
XX      Disclosure: Page 12-13; 28pp; English.
XX      The sequence is that of the promoter sequence of the endothelial
XX      protein C receptor (EPCR) gene. Such a regulatory element is
XX      useful to control expression of a gene/biologically active
XX      nucleotide molecule, by expressing these under control of one
XX      of the elements (optionally with the thrombin response element).
XX      Expression of the gene/nucleotide molecule is selectively in
XX      large vessel endothelial cells and/or as a result of environmental
XX      stimuli (either thrombin or serum) can be achieved by inclusion of
XX      the appropriate regulatory element(s). Atherosclerosis and
XX      most other vascular diseases primarily occur in large vessels,
XX      and for gene therapy for such diseases it is desirable to target
XX      endothelial cells, the primary defence mechanism against cellular
XX      infiltration and thrombosis. The constructs are therefore particularly
XX      useful in gene therapy, especially when the gene encodes a protein,
XX      or the nucleotide molecules are antisense, triplex forming, ribozymes
XX      or guide sequences for RNase P which are used to mutate or stop
XX      transcription of a particular gene. Such genes/nucleotide molecules
XX      may be expressed in vivo in patients or in cell culture. For example,
XX      endothelial response elements may be used for any gene therapy where
XX      systemic distribution is required, whilst large vessel endothelial
XX      cell response elements are useful for expression of thrombomodulin
XX      in large vessel endothelium to decrease clot propensity at atheromas
XX      or in autoimmune diseases; the environmental stimuli response
XX      element(s) are useful e.g. to deliver agents whose expression
XX      should be increased during increased thrombin/platelet activation
XX      or regional trauma. The regulatory elements are also useful as
XX      hybridisation probes, in increasing expression of recombinant
XX      proteins by exposure of the encoding construct to thrombin
XX      and in drug screening and design.
XX      Sequence 3360 BP; 912 A; 777 C; 670 G; 1001 T; 0 other;
XX      Query Match 1.6%; Score 140.4; DB 19; Length 3360;
XX      Best Local Similarity 74.8%; Pred. No. 9.1e-24;
XX      Matches 202; Conservative 0; Mismatches 66; Indels 2; Gaps 2;

```

